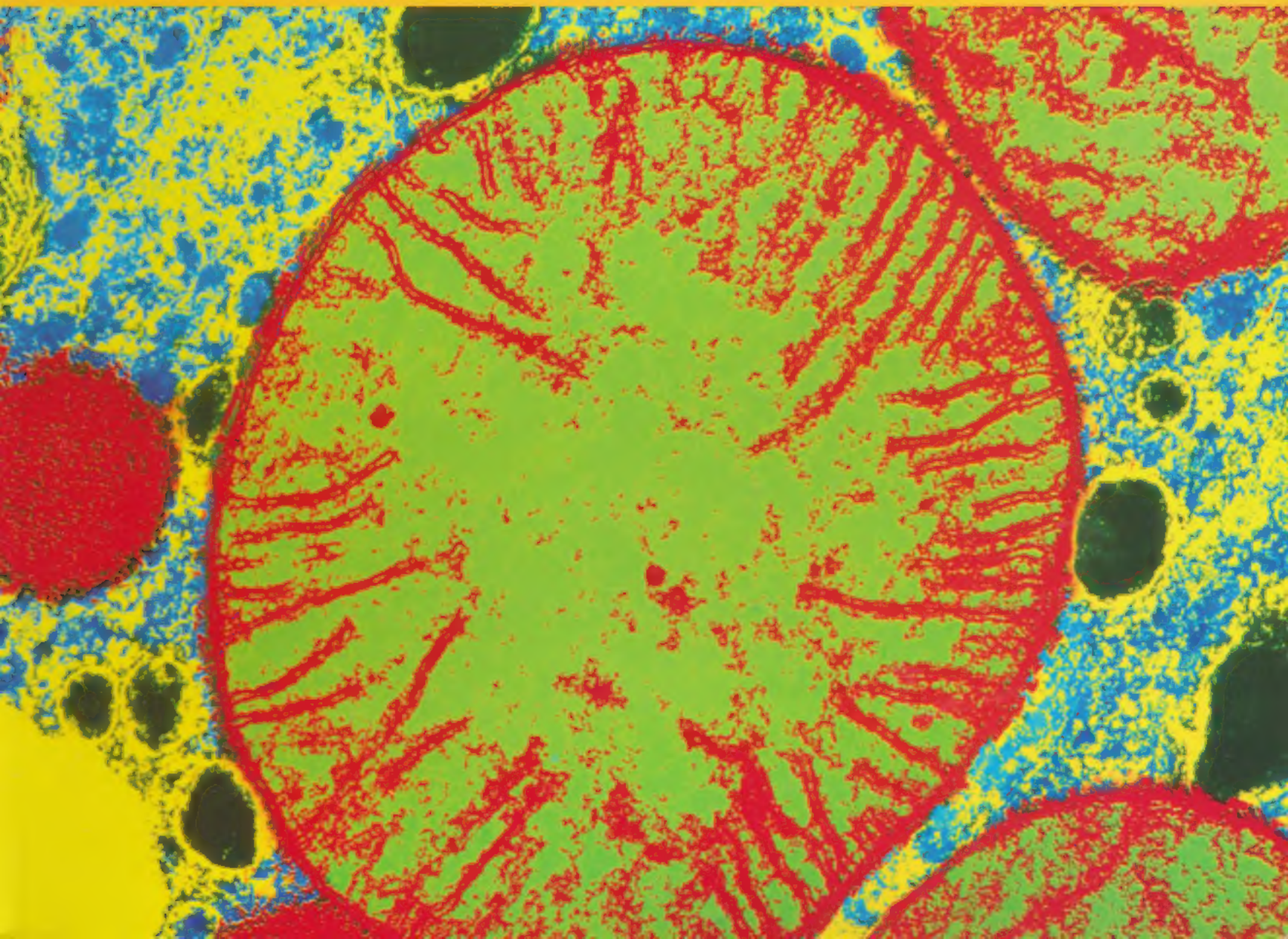
S204 Biology: uniformity and diversity
Science: Level 2

The Open University

# Generating Diversity

Edited by Michael Gillman



S204 Biology: uniformity and diversity
Science: Level 2

The Open University

# Generating Diversity

Edited by Michael Gillman

This publication forms part of an Open University course S204 *Biology: uniformity and diversity*. The complete list of texts which make up this course can be found at the back. Details of this and other Open University courses can be obtained from the Student Registration and Enquiry Service, The Open University, PO Box 197, Milton Keynes MK7 6BJ, United Kingdom: tel. +44 (0)845 300 60 90, email general-enquiries@open.ac.uk

Alternatively, you may visit the Open University website at http://www.open.ac.uk where you can learn more about the wide range of courses and packs offered at all levels by The Open University.

To purchase a selection of Open University course materials visit http://www.ouw.co.uk, or contact Open University Worldwide, Walton Hall, Milton Keynes MK7 6AA, United Kingdom for a brochure. tel. +44 (0)1908 858793; fax +44 (0)1908 858787; email ouw-customer-services@open.ac.uk

The Open University
Walton Hall, Milton Keynes
MK7 6AA

First published 2001. Second edition 2007



Edited and designed by The Open University.

Typeset by The Open University.

Printed in the United Kingdom by Latimer Trend and Company Ltd, Plymouth.

ISBN 978 0 7492 1449 4

2.1


The paper used in this publication contains pulp sourced from forests independently certified to the Forest Stewardship Council (FSC) principles and criteria. Chain of custody certification allows the pulp from these forests to be tracked to the end use (see www.fsc.org).

# The S204 Course Team

### Course Team Chair
Hilary MacQueen

### Maintenance Course Team Chair
Colin Walker

### Academic Editor
Caroline Pond

### Course Managers
Vivien Bacigalupo
Chris Edwards
Alastair Ewing
Christine Gardener

### Course Secretary
Dawn Partner

### Authors
Mary Bell (Book 5)
Eric Bowers (Book 6)
John Burnett (Book 4)
Alan Cann (Book 4)
Melanie Clements (Book 3)
Basiro Davey (Book 2)
Hilary Denny (Book 5)
Sue Downs (Book 5)
Mandy Dyson (Books 2 and 6)
Anna Furth (Books 3 and 4)
Michael Gillman (Book 2)
Tim Halliday (Book 2)
Jane Loughlin (Book 3)
David Male (Book 4)
Hilary MacQueen (Book 4)
Judith Metcalfe (Book 3)
Paul O'Shea (Book 3)
Phil Parker (Book 5)
Caroline Pond (Books 1, 2 and 6)
Irene Ridge (Books 1, 4 and 5)
Jerry Roberts (Book 5)
David Robinson (Book 6)
Jill Saffrey (Book 3)
Robert Saunders (Book 6)
Ayona Silva-Fletcher (Book 3)
Valda Stevens (Book 2)
Margaret Swithenby (Books 3 and 4)
Colin Walker (Books 3 and 4)

### GLO Editor
Peggy Varley

### Editors
Pat Forster
Ian Nuttall
Gillian Riley
Bina Sharma
Margaret Swithenby
Peter Twomey

### OU Graphic Design
Sue Dobson
Ruth Drage
Sara Hack
Jenny Nockles
Pam Owen
Howard Twiner
Andrew Whitehead
Liz Yeomans

### Library
Lydia Eaton
Judy Thomas

### Picture Research
Lydia Eaton

### Book Assessors
Gianfranco Novarino (Book 1)
Aubrey Manning (Book 2)
Karl Swann and David Harris (Book 3)
Peter White (Book 4)
David Clarkson and Rachel Leech (Book 5)
John Currey and Brian James (Book 6)

### External Course Assessor
Professor Sir David Smith

### Consultants
Patricia Ash
Sue Downs
Christine Gardener
Jean Macqueen
Terry Whatson

### Skills
Patricia Ash
Hilary Denny

## CD-ROM and Video Production

Mary Bell
Gail Block
Jacki Brown
Phil Butcher
Hilary Denny
Mike Dodd
Susan Dugher
Anna Furth
Phil Gauron
Michael Gillman
Tim Halliday
Nicky Heath
Caryl Hooper
G. D. Jayalakshmi
Juliet Kauffmann
Martin Kemp
William Kirk
Hilary MacQueen
Derek Martin
Mark Murphy
Phil Parker
Caroline Pond
Brian Richardson
Irene Ridge
Jerry Roberts
David Robinson
Jill Saffrey
Liz Sugden
Andrew Sutton
Margaret Swithenby
Gary Tucknott
Verina Waights
Darren Wycherley

## Photo on cover

### Mitochondria

False-colour transmission electron micrograph of mitochondria (green) shown in cross-section.

In eukaryotic organisms mitochondria are the sites of respiration, the chemical process that uses molecular oxygen to oxidise sugars and fats to release energy. The energy is stored as adenosine triphosphate (ATP) and is used by the cell to drive chemical reactions such as protein synthesis. Mitochondria are bound by a double membrane; the inner membrane is folded to produce ingrowths (red lines) called cristae, which are where the chemical reactions of respiration occur.

Mitochondria occur in virtually all eukaryotic cells, although this particular example is taken from a mammal. In contrast, bacterial cells (prokaryotic) do not contain mitochondria. However, mitochondria are widely believed to have originated as free-living prokaryotic cells that were 'captured' within eukaryotic cells. Consequently there are many structural and functional similarities between mitochondria and bacterial cells. Mitochondria therefore illustrate the theme of uniformity across the diverse range of living organisms: plants, animals and microbes.

The micrograph also includes other cellular organelles, visible in the surrounding cytosol.

Courtesy of Dr Gopal Murti/Science Photo Library.

# CONTENTS

# SURVIVING THE WINTER

## 1.1 LIVING IN A FLUCTUATING ENVIRONMENT

The great majority of organisms live in environments that fluctuate extensively and cyclically. For example, an animal such as a limpet living on a sea-shore experiences a 12-hour cycle in which it is alternately exposed, first to the air, the wind and the heat of the Sun, then to pounding by the sea. Life in the natural world exposes organisms to a diversity of environmental cycles, acting on a variety of time-scales. The most obvious cycles are night and day, with a period of 24 hours, and winter and summer or wet and dry seasons over a year; less obvious, but important for many organisms, is the lunar cycle (28 days). Even organisms that live deep in the oceans or under the ground may be exposed to cyclic variations in their environment.

In this chapter, we study one aspect of the fluctuating nature of an organism's environment. We consider how organisms living in a temperate climate, such as that in Britain, are adapted to cope with winter. You will see that there is much diversity of adaptations among organisms, with different species coping with the demands of a fluctuating environment in quite different ways. As cyclic variations are a widespread feature of environments, the range of adaptations to them is an important source of biological diversity.

### 1.1.1 A DIVERSITY OF STRATEGIES

Faced with an environment that becomes relatively hostile with the onset of winter, an adult organism can, broadly speaking, do one of four things:

1 It can maintain an active lifestyle, adapting in various ways to the prevailing conditions. The robin (*Erithacus rubecula*) is an example of such a species; so are evergreen trees.

2 It can abandon an active lifestyle, adopting an inactive existence for the duration of winter. The hedgehog (*Erinaceus europaeus*) spends the winter in hibernation. Many plants remain dormant below ground during the winter.

3 It can die before the onset of winter, as many adult insects do, but their offspring can survive the winter, for example as an egg or a pupa. Some plants survive the winter as seeds.

4 It can leave, migrating to a part of the world where conditions are favourable during the winter months. Each year, swallows (*Hirundo rustica*) leave Britain in autumn and migrate to Africa, whilst other birds such as barnacle and pink-footed geese migrate to Britain from further north.

Each of these four strategies is more applicable to some kinds of organisms than others. Long-distance migration, for example, is only an option for species that can store energy or can feed continuously (e.g. reindeer, *Rangifer tarandus*). For each of the three groups of species in Table 1.1 there is one strategy that is not an

option. Later in this chapter, we will focus on each strategy in turn, examining how it is manifested in different groups of organisms. Table 1.1 summarizes these alternative strategies and gives some examples of organisms that use each of them. This table provides a conceptual framework for much of the chapter.

**Table 1.1** A framework for considering the diverse ways in which selected groups of organisms have adapted to winter in a temperate region, such as Britain.

| Strategy | Plants | Insects | Vertebrates |
|---|---|---|---|
| 1 *'Tough it out'*<br>Maintain an active adult existence by altering behaviour and physiology with changing conditions. | Plants that continue to photosynthesize, e.g. evergreen trees | Not an option for most insects | Mammals that remain active, e.g. fox<br>Birds that remain active, e.g. robin |
| 2 *'Opt out'*<br>Maintain an inactive existence as adults for the duration of the hostile period. | Perennial plants that survive the winter by dying down or going dormant above-ground and/or forming subterranean storage organs, e.g. bulbs and rhizomes | Survive the hostile period in a state of torpor, e.g. certain wasps, bees and butterflies. | True hibernation in small mammals<br>Not an option for most birds<br>Winter torpor in amphibians and reptiles<br>Freeze-tolerance in amphibians |
| 3 *Juvenile survival*<br>Survive the hostile period as some non-adult phase of the life cycle. | Survive the hostile period only in the form of seeds. | Survive the hostile period in the form of eggs, larvae or pupae. | Not an option for vertebrates |
| 4 *'Go away'*<br>Migrate as adults to a location where conditions are favourable. | Not an option for plants | Some butterflies migrate north each spring and summer. | Migration in some birds and mammals |

Each of the four strategies can be discussed at different *levels of explanation*. For example, with migration of birds we can consider the molecular and cellular processes involved in energy storage and its utilization for long-distance flight. We can discuss the physiological and structural features of birds' wings that allow long distance flight. Alternatively, we can consider migration in an evolutionary context, discussing the selection pressures that favour migration. This approach leads us into the ideas of *types* of explanation. We will see firstly that organisms carry out certain activities, such as breeding, at those times of year when it is most adaptive for them to do so (i.e. to maximize their fitness, detailed further in Chapter 3) and, secondly, that those activities are performed in response to appropriate environmental cues. For example, many animals whose food is available mainly in summer ensure their survival in winter by building up fat reserves in the autumn.

It is important to separate arguments about adaptation during evolution, referred to as *ultimate* types of explanation, from the analysis of causal mechanisms, referred to as *proximate* types of explanation. To this end, a distinction is made between proximate and ultimate factors. **Ultimate factors** are those features of the environment that have favoured the evolution of particular adaptations: for example, many small mammals have evolved hibernation as an adaptation for coping with adverse winter conditions. **Proximate factors** are those specific features of the environment that elicit specific responses by organisms: long nights elicit a variety of responses in many small mammals, such as increased foraging, laying down fat and hoarding food. Thus proximate factors may operate at several levels. We will see that uniformity and diversity can only be fully understood with reference to both types of factors (ultimate and proximate) as well as several different levels of explanation.

## 1.2 RESPONSE TO WINTER: UNDERSTANDING AT DIFFERENT LEVELS

Winter in a temperate region poses a number of environmental problems for organisms. Most obviously, average temperatures are lower than at other times of year and there are frequent frosts. Frost is highly significant for living organisms because water forms such a large proportion of their body tissues: for the great majority of organisms, freezing of their tissues leads to death. Secondly, because, as shown in Table 1.1, many adult organisms die, go into hiding or migrate in winter, many of those animals that remain have a greatly reduced supply of food (i.e. plants for herbivores and prey for carnivores). A third significant factor in winter is that the daylight period becomes markedly shorter than at other times of the year. For plants that require light for photosynthesis and for animals that use daylight vision to find food, shorter days in winter severely restrict their ability to acquire the energy and nutrients that they need to support life. Finally, a major problem for plants in winter is shortage of water, which arises for two reasons. Firstly, when water becomes frozen in the ground, it is no longer available for plants to take up through their roots. Secondly, low temperatures greatly reduce the capacity of plant roots to take up water that is available, because their physiological processes are slowed down and, often, the transport tubes (xylem) become blocked by air bubbles, so are unable to function. Here we will introduce these problems at a number of levels: Sections 1.2.1–1.2.3 invoke proximate types of explanation whilst Section 1.2.4 considers ultimate types of explanation.

### 1.2.1 THE MOLECULAR LEVEL

It is common knowledge that the freezing point of pure water is 0 °C. Often, however, the temperature of water can fall below 0 °C without it freezing, for two reasons:

1 Any solvent containing a dissolved substance has a lower freezing point than when pure, which is why the sea freezes at a lower temperature than clean freshwater.

2 The occurrence of **supercooling**, the phenomenon by which a fluid remains liquid at a temperature below its normal freezing point. Freezing occurs when

water molecules become aligned in a particular pattern that leads to the formation of ice crystals. It begins when two molecules become aligned, a process called nucleation, and ice grows from a *nucleation point*. Whether or not water molecules become appropriately aligned depends on the conditions, so the temperature at which freezing occurs varies within a narrow range. Supercooled water freezes abruptly if an ice crystal is added to it. Small objects such as microbes also act as nucleation points, including certain bacteria on the surfaces of plant leaves. For example, *Pseudomonas syringae* cells act as ice-nucleation points, causing ice to start forming at −1 °C. In the absence of these bacteria, ice does not form on plants until about −5 °C. Genetically engineered strains of *P. syringae*, called 'ice-minus', have been produced which lack ice-nucleating characteristics. When these bacteria are sprayed onto the leaves of crop plants, they make them more frost-resistant

Fishes living in temperate and arctic habitats remain active through the winter (strategy 1 in Table 1.1), swimming about in very cold water. The winter flounder (*Pseudopleuronectes americanus*), for example, swims actively in water at a temperature well below the freezing point of its blood. What prevents its tissues from freezing? Two hypotheses, both of which invoke molecular levels of explanation, have been proposed. The ice-prevention hypothesis is that the tissues of fishes contain compounds that act as antifreeze, lowering the overall freezing point of their cells and body fluids. A number of such compounds have been identified, but they do not occur in some body fluids, such as urine and the fluid in the eye, suggesting that freezing is only prevented in some parts of the body. The alternative hypothesis, proposed by Valerio *et al.* (1992) is that ice is excluded from the body by a surface barrier. The skin of the winter flounder contains peptides (short chains of amino acids) which prevent ice formation. These molecules act in two ways: by lowering the freezing point of water in the skin and by binding to water molecules and so preventing them from binding to each other to form ice crystals. As described above, ice grows from nucleation points and the skin prevents ice formation at the fish's interface with very cold water.

### 1.2.2 THE CELLULAR LEVEL

The water in plant and animal tissues has two major components: the intracellular fluid (within cells) and the extracellular fluid, which fills the spaces between cells. When a tissue freezes, ice typically forms first in the extracellular fluid. Ice formation has two harmful effects:

1. It disrupts cell walls and cellular membranes.
2. The formation of ice in extracellular fluid effectively removes water from solution, thereby increasing its solute concentration. The gradient in solute concentration between the extracellular fluid and the cells it surrounds causes water to move out of the cells into the extracellular spaces (a process called osmosis), so that the cells collapse.

The destructive effect of frost on plants is a familiar consequence of these forms of cellular dysfunction. Frosted foliage collapses and turns brown, a result of the destruction of tissues in which ice has formed. These effects are especially obvious in some non-native plants, which are not adapted to survive frost.

## 1.2.3 THE PHYSIOLOGICAL AND BEHAVIOURAL LEVELS

### ANTICIPATION OF WINTER

If organisms are to survive winter, they must be well prepared for cold weather and sometimes for reduced supply of food. Physiological changes such as shedding leaves and building up fat reserves, and behaviours such as hoarding food, must be completed before winter begins. In order to be prepared, organisms need to anticipate the onset of winter, which is done in two ways, each of which has its own associated physiological mechanisms. First, they could respond to environmental cues that predict that winter is imminent, e.g. lower temperatures and shorter day lengths.

❍ Which of these cues is the more reliable?

● Changes in day length, which follow an identical pattern every year. We know from personal experience that temperature is very variable and that cold weather comes earlier in some years than others.

Secondly, animals could have an internal clock which tells them what time of year it is, just as a calendar tells us the date.

### PHOTOPERIODISM

A great deal of research has been carried out into the way that organisms respond to changes in day length, i.e. the relative durations of light and darkness in a 24-hour period. This relationship is expressed as the ratio of the number of hours of light (L) to the number of hours of darkness (D), i.e. L : D. **Photoperiodism** is defined as the responses of organisms to changes in the L : D ratio. For example, flowering plants can be divided into three categories on the basis of the effect of photoperiodism on the onset of flowering:

- *Short-day plants* flower in early spring (e.g. primrose, *Primula vulgaris*) or in the autumn (e.g. *Chrysanthemum* spp.), when L is small relative to D.
- *Long-day plants* flower in the summer (e.g. potato, *Solanum tuberosum*, or lettuce, *Lactuca sativa*), when L is large relative to D.
- *Day-neutral plants* are not affected, in terms of flowering, by changes in the L : D ratio (e.g. groundsel, *Senecio vulgaris*).

Horticulturalists have long been familiar with these effects and exploit them to get plants to flower at times that suit them. If you go to the Chelsea Flower Show, which takes place in May, you see spring, summer and autumn plants, raised under artificial conditions, all in bloom at the same time.

❍ How can chrysanthemums be made to flower later in the year than normal?

● By keeping them under high L : D conditions during the spring, using artificial lights, and transferring them to low L : D conditions just before they are required to bloom.

Simply observing that primroses, for example, flower in early spring is consistent with the hypothesis that they flower in response to short light periods, but it does not preclude alternative hypotheses, such as that they flower in response to increased temperature. However, primroses kept indoors under consistently long light periods and a variety of temperatures do not flower, confirming that onset of flowering is indeed a response to short light periods. Photoperiodism can only be revealed by means of experimental manipulations in which plants are kept under different L : D regimes, and in which other environmental cues to which they might respond are held constant. Variations on this kind of experiment tell us other things about photoperiodism. For example, some plants flower in response to a single exposure to an appropriate L : D ratio. Others flower only if an appropriate L : D ratio is sustained for several days. A pigment, called phytochrome, which switches between 'dark' and 'light' forms, is involved in sensing the length of the light and dark periods.

For a specific physiological response, there is typically a particular L : D ratio at which the response starts to occur. The **critical photoperiod** is defined as that L : D ratio at which 50% of the population being studied switches from one state to another. There is considerable variation in the preciseness of a particular organism's response to a critical photoperiod. Figure 1.1 shows two examples of critical photoperiods for animals. Whereas *all* aphids switch from sexual to asexual reproduction at an L : D ratio of 10 : 14, male sparrows start to show testicular development at an L : D ratio of about 8 : 16, but all do not mature unless the ratio is about 16 : 8.

In some animals, the physiological basis of this effect is known. The hormone melatonin is secreted by the brain during the dark period, with the result that in winter blood levels of melatonin are much higher.

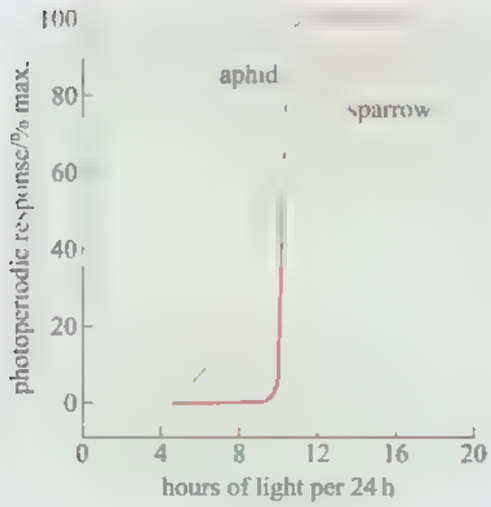



**Figure 1.1** Critical photoperiodic responses in two species: transition from sexual to asexual reproduction in the vetch aphid (*Megoura viciae*) and testicular development in the white-crowned sparrow (*Zonotrichia leucophrys*).

### CIRCANNUAL CLOCKS

The physiology and behaviour of microbes, plants and animals show cyclical changes on a diversity of time-scales. Most familiar to humans are circadian rhythms, which determine patterns of sleep and wakefulness and of changes in body temperature over a 24-hour period. There is also evidence that certain annual rhythms are controlled by internal clock mechanisms, called **circannual clocks** (circannual means 'about a year').

Whatever the time-scale over which they operate, the existence of endogenous clocks can only be determined by experimental investigation. Simply observing that an animal or a plant shows a daily or an annual cycle of activity or physiology does not preclude the possibility that it is responding to rhythmic changes in the environment. The crucial experiment is called a **free-running** experiment, in which organisms are kept under conditions in which they are unable to detect normal cycles in external features of their environment. For example, captive alpine marmots (*Marmota marmota*, a member of the squirrel family) have been kept for long periods under constant temperatures and constant L : D ratios. Despite having no known cues that winter was approaching, they showed a 100% increase in food intake in the autumn, just as they do in nature, in preparation for hibernation. Kenagy (1981) found that chipmunks (*Eutamias minimus* and *E. amoenus*, also in the squirrel family) kept under combinations of three photoperiods (L : D ratios 8 : 16, 12 : 12 and 16 : 8) and two temperatures (5 and 23 °C) maintained normal cyclical patterns of testis growth, body mass, water consumption, locomotion and winter torpor.

### LINKAGE OF ANIMAL REPRODUCTIVE CYCLES TO WINTER

The reproductive success of organisms is crucially determined by the time of year when they breed. Many birds living in Britain breed in the spring, with the result that they are able to feed their young at that time of year when there is most suitable food available. For many kinds of animal, all parts of the reproductive process, from mating to birth, follow the end of the winter. A complication for some larger mammals, however, is that there is a long gestation period (the interval between conception and birth, nine months in humans). A long gestation period and giving birth early in the spring are not easily reconciled with being inactive in winter. Figure 1.2 shows how the reproductive cycles of four mammals native to Britain are related to the winter.

The hedgehog (*Erinaceus europaeus*) has a short reproductive cycle and so can complete the entire process in the spring, mating soon after emerging from hibernation. The red deer (*Cervus elaphus*), with a long gestation period, mates in the autumn (the rut) and gives birth in spring, remaining active through the winter. The other two species shown in Figure 1.2 have gestation periods that are too long for the entire reproductive cycle to be completed in the spring, but too short to occupy the whole winter. Badgers (*Meles meles*) mate in the autumn and the eggs are fertilized immediately. Implantation of the zygote into the wall of the uterus is, however, delayed for several months, during which the female spends short cold spells in a state of torpor in an underground den. The noctule bat (*Nyctalus noctula*) mates in the autumn but the eggs are not fertilized. Instead,


**Figure 1.2** Reproductive cycles of (a) hedgehog, (b) red deer, (c) badger and (d) noctule bat, in relation to the winter. M = mating; F = fertilization of the egg(s); G = gestation; B = birth. In the badger, implantation occurs 3–9 months after fertilization. In the noctule bat, fertilization is delayed for up to 7 months following mating.

females store sperm in their reproductive tract until late winter, when fertilization occurs and gestation begins. Throughout this time, they are hibernating.

These different reproductive patterns are controlled by environmental cues in much the same way as the flowering of plants. Red deer, badgers and bats are called *short-day breeders*, the development of their gonads and their sexual behaviour is stimulated by the lengthening dark period characteristic of autumn. Hedgehogs and many other small mammals are *long-day breeders*; their gonad development and sexual behaviour are triggered by a decrease in the dark period.

Figure 1.2 illustrates how diverse are the reproductive cycles of a single group of animals, the mammals. Deer, badgers and hedgehogs are very different animals, however, and before we leave this brief account of seasonal breeding, it is important to emphasize that there is enormous diversity, even among closely related species. Figure 1.3 shows the breeding seasons of members of a single genus, *Peromyscus*, in North America. *Peromyscus* is a genus of small American rodents that includes a variety of mice, such as the deermouse, white-footed mouse and cactus mouse.

) How does the breeding season in *Peromyscus* species change with latitude?

● The breeding season of *Peromyscus* is limited to a single, three-month period in northerly latitudes, but is continuous through the year in Mexico. In between, there is considerable variation in the duration of the breeding season; in some regions at intermediate latitudes, there are two peaks in breeding activity.

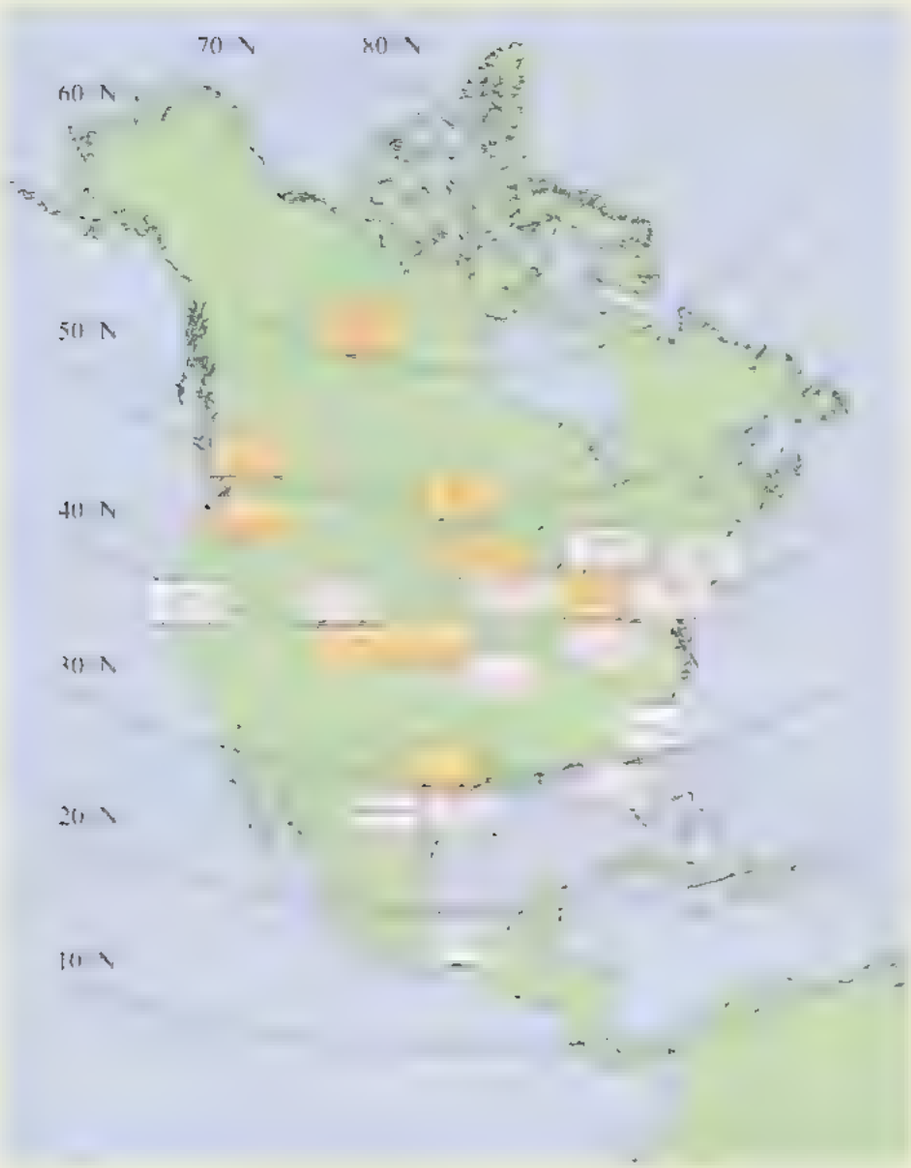



**Figure 1.3** The breeding seasons (expressed as the relative proportion of females pregnant in each month) of several populations of *Peromyscus* species in North America. Each box represents a 12-month period from January (left) to December (right). The yellow boxes are all for a single species, the deermouse (*P. maniculatus*). Data from Bronson (1987).

### WINTER FAT RESERVES AND REPRODUCTION

For some animals in which breeding begins as soon as the winter has finished, energy reserves accumulated in the previous autumn may be important, not only for winter survival but also for reproductive success in the following spring. In newts, for example, as in other temperate amphibians, nutrient reserves are built up in the late summer and autumn, in both the liver and the abdominal fat stores. These reserves are only partly used up in the winter, when newts are largely torpid underground, and the remainder plays a key role in reproduction. Female newts use their fat reserves to produce yolk for their eggs; males use theirs to develop a large dorsal crest, which is crucial in mating, females responding positively only to males with large crests (Figure 1.4). A population of great crested newts (*Triturus cristatus*) on the Open University campus has been studied in detail by John Baker. He weighed and measured males as they

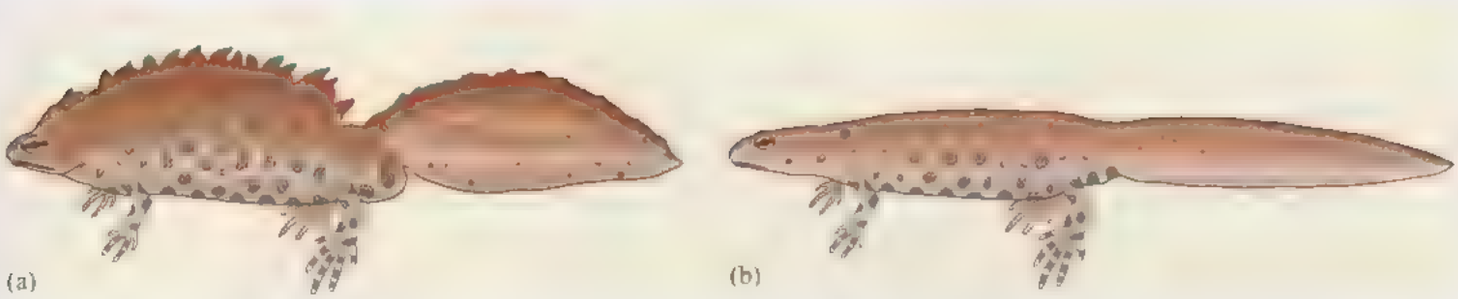



**Figure 1.4** (a) Male and (b) female great crested newts (*Triturus cristatus*) in breeding condition.

migrated towards their breeding pond in early spring and, for each male, he calculated a 'condition index'. Males in good condition (i.e. with a condition index value greater than zero) were heavy for their length, because of their larger fat reserves. Later in the year, Baker recaptured the same individuals, now in the pond, and measured their dorsal crest, which develops after newts have entered the water. He found a positive correlation between their crest height and their condition index measured a few weeks earlier (Figure 1.5). Thus, a male newt's attractiveness to females, and hence his reproductive success, is partly determined by the amount of fat that he has left over after hibernating during the winter.

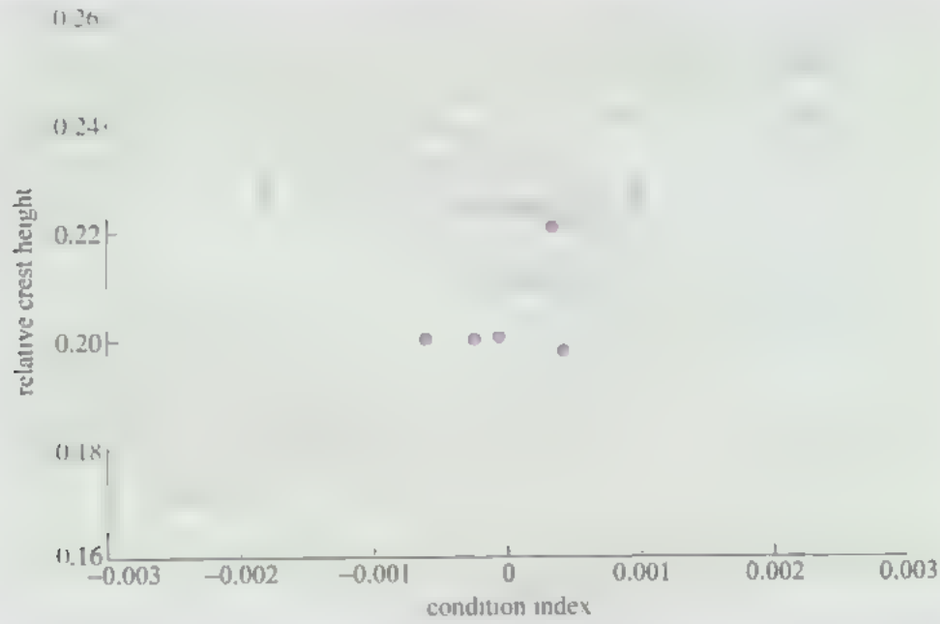



**Figure 1.5** The relationship between a male great crested newt's condition index when he arrives at a pond and the size of the dorsal crest which he develops a few weeks later. Data from Baker (1992).

## 1.2.4 THE EVOLUTIONARY LEVEL

### LIFE HISTORIES AND TRADE-OFFS

In this section, the emphasis switches from proximate (molecular, cellular, physiological and behavioural) types of explanation to ultimate types of explanation. In order to proceed, we need to understand two key concepts: **life history** and **trade-off**. Both of these concepts are important tools in organizing thoughts about why organisms are so diverse. An organism's life history is the set

of key biological events in its life, including birth, growth, reproduction, sometimes migration, and death. Life histories are distinguished from life cycles, which are a detailed description of the sequence of stages in an organism's life. Life histories can be subjected to quantitative analysis. A familiar example of a life history is that of an annual plant which, following birth (germination), grows, flowers (reproduces), sets many seed and then dies. A contrasting life history is that of large-bodied mammals such as African elephants or blue whales. Following birth, these animals grow to a very large size before giving birth to (usually) one offspring at a time. There is then a period of recovery before new young can be produced. Death may follow 20–30 years after the first birth, during which time perhaps 8–10 young may have been born.

Between the extremes of an elephant and an annual plant lie the life histories of many species. If these life histories are subjected to quantitative analysis, for example by plotting fecundity (number of offspring) against chances of survival, the results indicate that trade-offs are at work. For instance, organisms with a high fecundity have a lower chance of future survival. What are the reasons for these results? The patterns between species reflect processes that have occurred within species and can be given proximate explanations. In proximate terms, each behavioural, physiological or biochemical activity has an energetic cost. For a given energy input (food or light intake) an organism cannot afford to indulge in many different costly activities. Therefore trade-offs have to occur. Let us take the example of a short-lived plant in which there is genetic variation in the amount of photosynthetic product (e.g. starch) that can be stored in the root. Some individuals can store more in the roots than others. Those individuals that store less in the roots divert the energy into seed production. The probability of plants surviving after flowering is inversely related to the amount of seed produced. Thus plants that put less stores into roots have a higher probability of dying after flowering, in contrast to plants that store more in the roots. The latter produce fewer or smaller seeds and hence have a higher probability of survival after flowering.

Natural selection can operate on this genetic variation. It may be that under certain environmental conditions the storers are favoured, and under other environmental conditions, those that produce many seeds are favoured. It may also be that both strategies are favoured by the same environmental conditions. Thus both seed producers and storers do well in winter, i.e. there are two equally adaptive solutions to the same environmental problem. Hence we can see how genetic variation underpinning life history variation due to trade-offs in individuals can, through natural selection and subsequent speciation, be represented as life history variation between species. (Genetic variation and natural selection are reviewed in Chapter 3.) In conclusion, it is possible to move from proximate explanations of trade-offs and life history variation in individuals, to ultimate explanations of life history variation between species. This book will cover many examples of trade-offs, which act as important constraints on organism structure.

### PLANT LIFE HISTORIES

To review ideas on life histories, let us consider some plant examples. Among plants, there are three broad categories of life history, each of which has different implications for how they respond to the onset of winter:

1 *Annuals* As described above, these plants complete their life cycle in a single year. Most annuals in temperate regions set seed during the summer and then die before winter (Table 1.1, strategy 3). They represent the extreme version of the high seed-producers discussed above.

2 *Biennials or short-lived herbaceous perennials* These plants have the capacity for storage of photosynthetic product(s). In this particular category are plants that live for two or a few years, flowering and setting seed only in the second or final year. Since biennials that did not survive their first winter would fail to reproduce, they must be adapted for winter survival. They allocate some of their resources to growth in the first year, and some to accumulating stored reserves in their roots. Their foliage may die back with the onset of winter (strategy 2) but their stored nutrients and energy enable them to grow quickly in the second year, prior to flowering. Thus, whereas growth (vegetative) and reproductive phases of the life history occur in the same year in annuals, they are separated into different years in biennials.

3 *Long-lived perennials* These plants may persist for many years, typically reproducing many times. Perennial plants include herbaceous plants that die back each year (strategy 2), and woody plants, which possess a number of adaptations for surviving the winter (strategies 1 and 2). There may be a long juvenile period before the first reproduction event.

For smaller biennials and perennials, the typical overwintering strategy is for those parts of the plant above ground to die back, leaving an underground storage organ, such as a tap root, a bulb or a rhizome, to survive the winter. For larger plants such as trees, this option is not viable, because reconstructing the above-ground parts of the plant every year would preclude any long-term growth. Trees have two main strategies for surviving the winter, belonging to either the **deciduous** or the **evergreen** category (strategies 2 and 1 respectively). Deciduous trees are those that shed all their leaves in a particular season, usually the autumn. Evergreen trees retain leaves all year round, they do drop and replace their leaves, but only some at a time. While trees do not possess discrete storage organs, they store energy-releasing compounds and other nutrients during the winter, in their trunks and roots.

Storage during the winter involves risks because, among temperate animals, there are herbivores that follow strategy 1, remaining active through the winter. Plant storage organs are a vital food source for such animals and many individual plants fail to survive the winter because their storage materials have been consumed by animals or fungi.

We will now discuss in more detail the means by which selected groups of organisms survive the winter, with reference to the four strategies in Table 1.1. Throughout the discussion of these strategies, we will move between different types and levels of explanation. It will also be clear that certain strategies are only open to certain taxonomic groups.

## SUMMARY OF SECTIONS 1.1 AND 1.2

1 The majority of organisms are exposed to environmental fluctuations, including seasonal change in climate. In this chapter, we focus on the effects of winter.

2 Organisms have evolved a range of strategies to cope with winter. Thus this common environmental variable has led to a diversity of responses.

3 The strategies for coping with winter can be considered with respect to different levels and types of explanation.

4 Molecular and cellular level responses to winter include the prevention of freezing through the production of antifreeze molecules such as peptides.

5 Physiological and behavioural responses include detecting the onset of winter through changes in the L : D ratio which prompts alteration of sexual behaviour. The reproductive cycles of many organisms are linked to the L : D ratio. Many of these effects can be investigated by experiment.

6 The life histories of organisms can be viewed as the products of trade-offs in biological processes.

## 1.3 STRATEGY 1: REMAINING ACTIVE THROUGH THE WINTER ('TOUGH IT OUT')

### 1.3.1 EVERGREEN PLANTS

In temperate regions, the most prominent evergreen plants are coniferous trees, or conifers (phylum Coniferophyta). Conifers dominate large portions of the Earth's land area, particularly at northern latitudes and high altitudes. This distribution reflects their ability to withstand long periods of cold weather. The major problem faced by conifers in winter is lack of water. Water that has turned into ice is not available to plants and, at freezing temperatures, plant roots are able to absorb such water as is available only very slowly. If conifers are not to die of desiccation they must, therefore, reduce the rate at which they lose water. The needle-like shape of conifer leaves reduces the rate at which water is lost from their surfaces and so reduces a tree's requirement for water.

Though very small compared to the leaves of most deciduous trees, pine needles are relatively thick, in comparison to the broad, flat leaves of deciduous trees. Consequently, their surface area is small relative to their volume, reducing water loss. Evaporation is further reduced by a thick, waxy cuticle that forms the outer surface of needles and by the stomata (pores through which leaves exchange gases with the air) being positioned in sunken pits.

Not all conifers are evergreens, and not all evergreens are conifers. There are some ten species of larches (genus *Larix*) that live mostly at high altitude in the Northern Hemisphere; all are deciduous, dropping their needles at the onset of winter. The holm oak (*Quercus ilex*), also known as the evergreen oak, is not a conifer, but retains its leaves through the winter. The holm oak is a native of continental Europe, from the Mediterranean to Brittany, that has been introduced into Britain.

As explained in Section 1.2.4, life histories involve trade-offs between many factors and this principle is well illustrated by a comparison between deciduous and evergreen trees. By retaining their leaves through the winter, evergreens do not bear the cost, as deciduous trees do, of reconstructing their entire photosynthetic apparatus each spring. However, the adaptations that enable their

leaves to survive the winter make them less efficient in the spring and summer than those of deciduous trees. Another trade-off is that conifers have a simpler system of water conducting cells which is less efficient when water is plentiful, but better (because it is less likely to block) when water is scarce and freezing occurs.

### 1.3.2 BIRDS AND MAMMALS

Birds and mammals are **endotherms**, meaning that they produce and retain a lot of heat within their own tissues, rather than absorb heat from their environment, as **ectotherms**, such as insects and reptiles, do. The terms endotherm and endothermy are now often used in preference to homeotherm and homeothermy, which refer to the ability of birds and mammals to maintain a more or less constant body temperature. Some endotherms, as you will see later, do not maintain a high body temperature at all times, and some ectotherms, such as larger reptiles, maintain a constant body temperature for long periods, even though the temperature of their environment changes.

Whilst endothermy allows some birds and mammals to remain active during winter, it also places formidable demands on those animals. Reduced environmental temperatures increase the amount of heat that they need to generate internally to maintain a constant temperature, at a time when the amount of food available to them is much reduced. For birds and many mammals, shorter day length in winter reduces the time available for finding food. It has been estimated that a small bird, such as a great tit, which feeds on seeds and small insects, has to find food items at an average rate of one every ten seconds during daylight hours through the winter. The extreme challenge posed to birds by the shortage of food in winter is illustrated by the impact of bird-tables in urban areas, which supplement their natural diet. Studies by the RSPB (Royal Society for the Protection of Birds) suggest that, in urban areas of Britain, the provision of food at bird-tables significantly increases the winter survival rate of several garden bird species.

#### MAINTAINING BODY HEAT

Mammals that remain active in winter maintain a core body temperature of around 37–38 °C; that of most birds is a little higher. In temperate habitats, thermal constancy is achieved despite the temperature of the environment varying between −20 and +20 °C. For animals in polar regions, the problem is even more severe, their environment varying between −60 and +20 °C. There are three principal ways in which endotherms can respond adaptively to cold conditions:

1. They can raise their metabolic rate so as to produce more heat to offset the increased heat loss.
2. They can improve external insulation between their body and the external environment, so as to reduce the rate of heat loss.
3. They can alter the pattern of blood circulation around their body so as to minimize the extent to which warm blood comes close to the skin (i.e. another form of insulation).

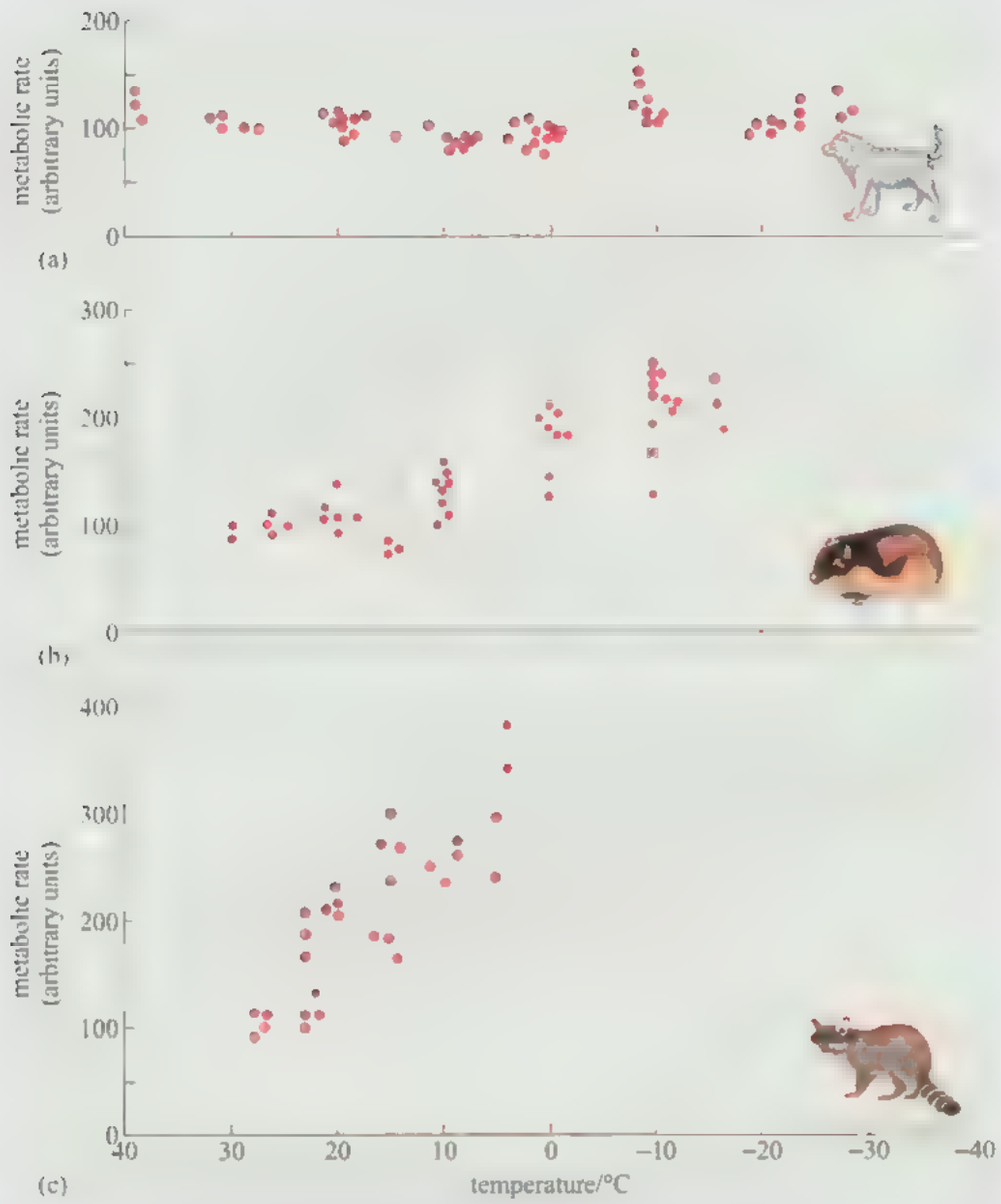



**Figure 1.6** The effect of declining temperatures on the metabolic rate of (a) an Eskimo dog, (b) an arctic lemming and (c) a tropical raccoon. Note that temperature decreases from left to right along the horizontal axis. Data from Irving (1966).

Considering the first of these processes, Figure 1.6 shows data from experiments in which three mammalian species, two from the Arctic and one from the tropics, were exposed to declining temperatures.

› In what way does the response to reduced temperature of (a) the Eskimo dog and (b) the arctic lemming differ from that of the tropical raccoon?

● (a) The metabolic rate of the Eskimo dog shows no clear pattern of change (except possibly increasing at very low temperatures), whereas that of the raccoon increases almost linearly as the temperature decreases from 28 to 5 °C. (b) The metabolic rate of the arctic lemming begins increasing at about 15 °C, but more gradually than that of the tropical raccoon.

Mammals and birds living at high latitudes generally rely on processes 2 and 3, rather than 1, though, as we have seen from Figure 1.6b, there are arctic species, such as the lemming, in which process 1 does play a major role.

Maintaining a high metabolic rate through a long winter requires that mammals and birds either maintain a high rate of food intake, or carry substantial energy reserves within their body, or show some combination of the two. They must also reduce heat loss during the winter. Many species maintain larger energy reserves, in the form of adipose tissue. The adipose tissue beneath the skin (subcutaneous fat) is also widely assumed to act as a thermal insulation layer, but this theory is erroneous. Adipose tissue is not much better an insulator than muscle. Only for marine mammals, such as seals, in which the adipose tissue forms a thick layer of blubber, is its role in thermal insulation significant. For birds and mammals, insulation is provided by feathers and fur respectively, which trap a layer of air next to the skin. Static air is a very poor conductor of heat, so that air trapped in plumage or fur reduces heat flow between an animal's skin and the outside. Bird plumage is a remarkably effective insulator: it represents only 5–7% of a bird's mass and trapped air makes up 95% of its total volume. The texture and often the colour of birds' plumage and mammals' fur changes with the onset of winter, in comparison to the summer. This change is effected during **moult**, a seasonal process in which a bird changes all or many of its feathers and a mammal replaces its summer coat with a thicker winter coat, and the converse in spring.

In birds, the plumage has two major components: contour feathers, which form the exterior surface; and down, which lies beneath the contour feathers. Most feathers consist, in varying proportions, of pennaceous (fan-like) and downy components (see Figure 1.7). The large feathers in the wings and tail are wholly pennaceous; many other small feathers close to the skin are pure down. In winter, the number of down feathers is markedly increased. Only in a few species has the total number of a bird's feathers been counted; humming-birds have about 940, whereas a swan has 25 000. In the house sparrow (*Passer domesticus*) there are 11.5% more feathers in winter than in summer. Each feather has an individual muscle which enables it to be lifted away from the body. This capacity enables birds to regulate heat loss by altering the thickness of the air layer trapped among the feathers.

**Figure 1.7** A body feather showing pennaceous (top) and downy (bottom) components.

In mammals, fur contains fine hair close to the skin, underneath thicker and larger surface hairs. The insulative effect of hair is determined primarily by its length; the longer the hairs, the better is the coat as an insulator (Figure 1.8). Comparison of the insulative value of the pelts of arctic and tropical mammals with hair of similar length reveals that the pelts of arctic species are only slightly better insulators (Figure 1.8). The difference arises because arctic species have more hairs per unit area of skin. Figure 1.8 also shows how the insulative value of the coats of two species, the beaver and the polar bear, is completely eliminated when it becomes thoroughly wet.

In many species, the winter plumage or coat is very different in colour and pattern from the summer one. This seasonal change has nothing to do with temperature control, but is usually related to camouflage. The winter plumage of many birds, especially males, is much less brightly coloured than the summer plumage. In many polar mammals, such as the arctic fox (*Alopex lagopus*), the winter coat is white, making the animal well camouflaged in snow.

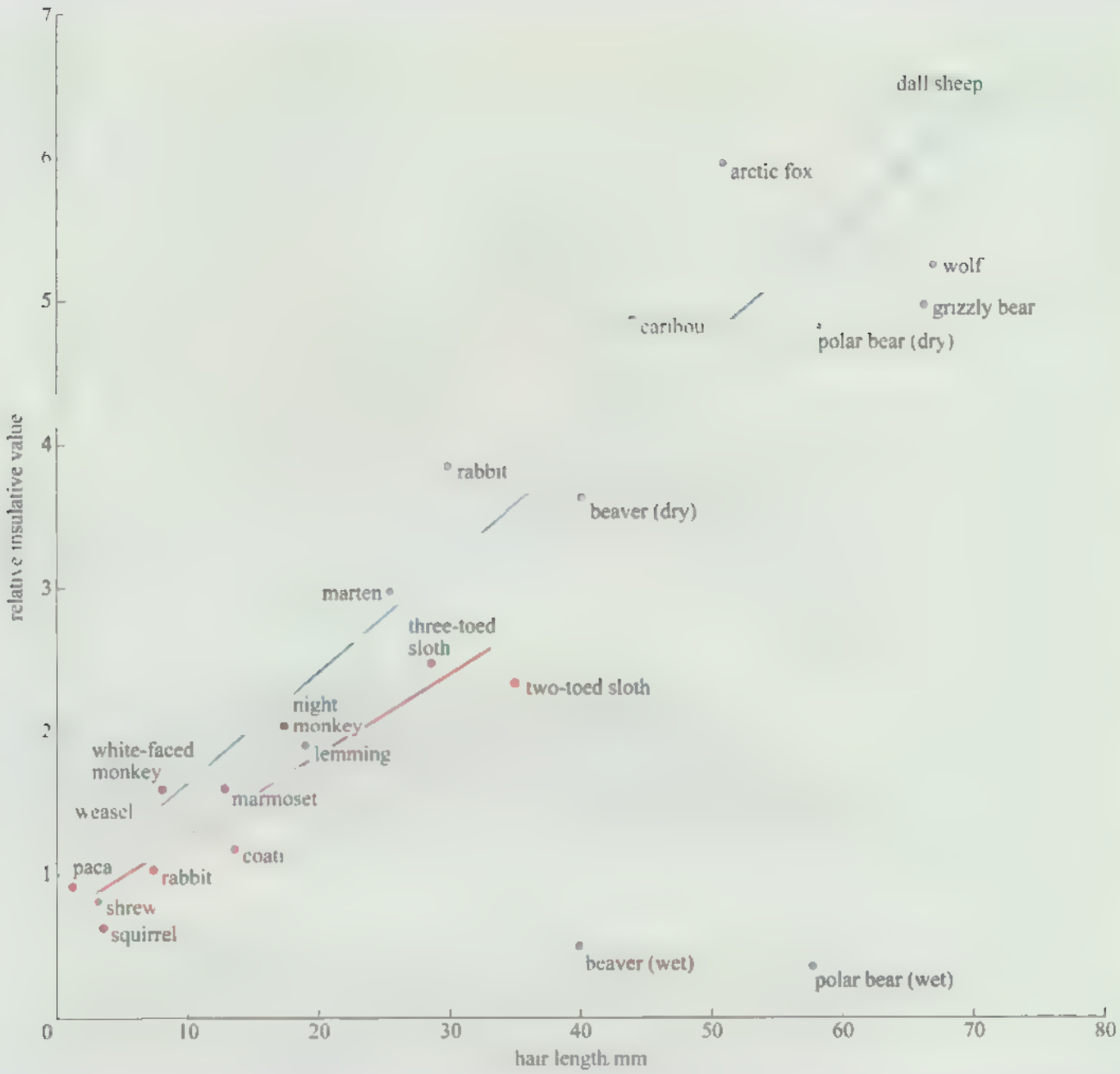



**Figure 1.8** Relationship between the insulative value of the coats of arctic (blue) and tropical (red) mammals and the length of their hair. This figure also shows the effect of wetting on a coat's insulative value. Data from Pough *et al.* (1996)

The third response to cold, varying blood flow to different parts of the body, is well illustrated by the Eskimo dog (Figure 1.9). At a temperature of −30 °C the Eskimo dog maintains a core temperature of around +38 °C, but adjustments in blood flow mean that the temperature of some parts of the body, notably the extremities such as the feet and the nose, is allowed to fall.

In many endotherms, cooling of the extremities is achieved by a heat exchange mechanism called a **rete mirabile** (pronounced 'reeta mirah-bilay' and meaning


**Figure 1.9** An Eskimo dog, at an environmental temperature of −30 °C, showing the temperature recorded at different parts of the body. Data from Irving (1966).

'wonderful net', Figure 1.10). Warm blood passing in arteries towards the skin runs close to colder blood passing back in veins towards the body core. Thus warm blood passing outwards gives up much of its heat before it reaches the skin. This effect is enhanced by constriction of blood capillaries near the skin when it is cold.

Very small mammals and birds that remain active in winter face a problem. Because their surface area is relatively large in relation to their mass, they lose heat in cold conditions faster per unit mass than larger animals and the amount of extra fur or feathers they would need to insulate themselves would make them immobile. The dwarf hamsters (*Phodopus* spp.) of Siberia and Mongolia have adapted to this problem in some interesting ways. They remain active in winter but run around, feeding on seeds and vegetation, under the deep snow, where they are not exposed to the wind and the extremes of cold. Whereas larger animals typically build up their adipose tissue and put on weight as winter approaches, dwarf hamsters get markedly lighter and leaner. Males lose about 50% and females 30% of their body mass and adipose tissue falls from 35 to 5% of the total body mass.

- How might a lower body mass enable a dwarf hamster to better survive the winter?

- Having a smaller body reduces an animal's energy requirement for locomotion.

Finally, consider the unfortunate pig. Pigs are large mammals and, as the result of selective breeding during domestication, they have so little hair that it provides no insulation against the cold. Pigs have the usual depots of subcutaneous fat found in other comparable mammals, but pig fat is no thicker than that of other mammals of similar size; indeed, domestic pigs have been selectively bred to be lean. The subcutaneous fat of pigs contributes little to their ability to survive cold weather. In cold conditions, the skin temperature of a pig falls much more than that of a human, due to restriction of the blood supply to the subcutaneous fat, so the poor hairless animal may feel cold in winter.

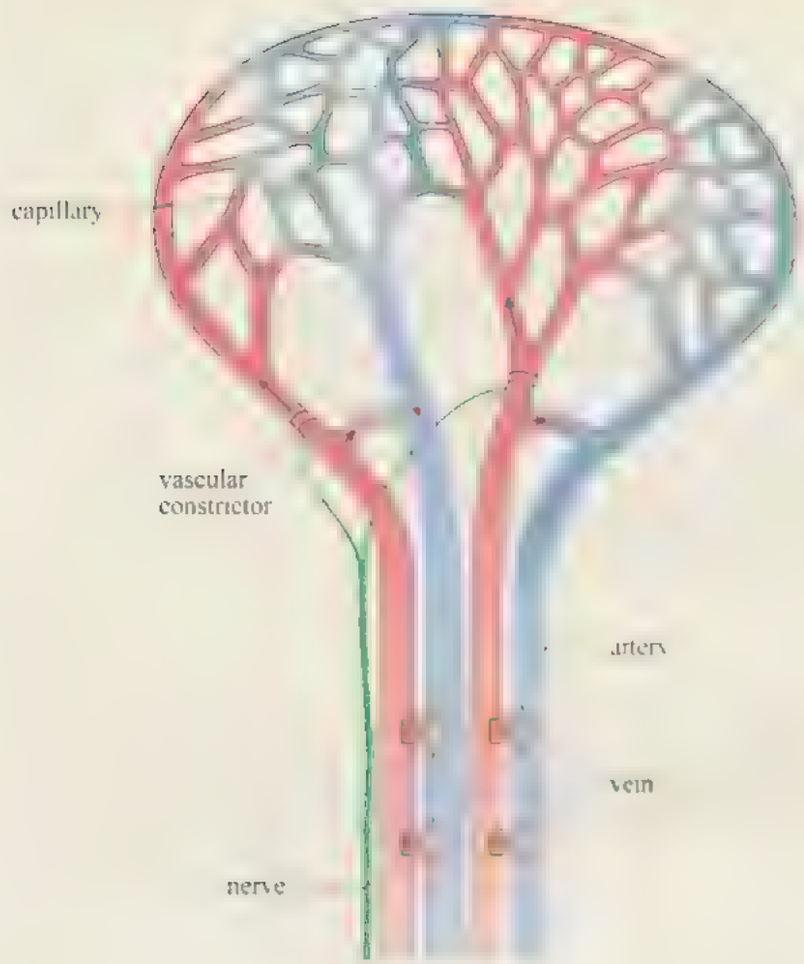



**Figure 1.10** Schematic diagram of heat exchange in a single rete mirabile under the skin of an endotherm. The small black arrows show the direction of blood flow and the broad orange arrows show the direction of heat flow. Blood flow from the small arteries into the capillaries is controlled by vascular constrictors.

### STORING FOOD IN WINTER

As described above, some animals solve the problem of remaining active in winter, when food is in short supply, by enlarging internal energy reserves in the adipose tissue. A problem with this strategy is that it increases body mass, which increases the energy cost of locomotion and, in extreme cases, may reduce mobility. An alternative strategy is to make external stores of food during the summer, when food is abundant, and to draw on them during the winter. Such behaviour, shown by a variety of birds and mammals, is called **hoarding** or **caching** (pronounced 'cashing')

There are two general kinds of hoarding behaviour. Larder hoarding involves building up one or a small number of large food stores. Scatter-hoarding consists of hiding many individual food items in separate places over a wide area. An essential feature of a food cache, from the perspective of the animal that made it, is that it must not be exploited by other individuals of the same or different species. Typically, animals prevent theft by active defence in the case of larders, or by concealment if their caches are scattered. Scatter-hoarders include crows, nuthatches, tits, jays, squirrels and foxes. larder-hoarders include many small mammals, some birds, and honeybees.

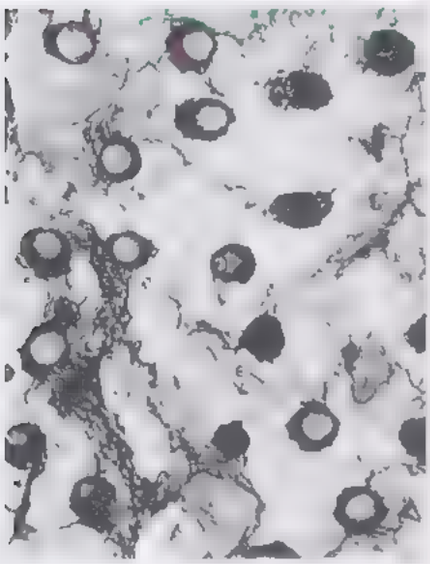


**Figure 1.11** Part of an acorn woodpecker communal larder. Numerous holes have been drilled into a tree, and many of them contain acorns.

Pikas (genus *Ochotona*) are herbivorous, guinea pig size animals, belonging to the same mammalian order (Lagomorpha) as rabbits and hares, which live in high altitude and desert habitats of Asia and North America. They do not hibernate in winter but rely on cached food to supplement what they can find to eat beneath the snow. In late summer they gather vegetation, storing it as a large haystack in a crevice beneath a rock, which they defend against their neighbours. Laboratory experiments on a variety of mammals reveal that hoarding behaviour is triggered by cooler temperatures and shorter days which are characteristic of autumn.

Among birds, an example of a larder-hoarder is the acorn woodpecker (*Melanerpes formicivorus*) of western North America. During the summer, acorn woodpeckers collect acorns and store them in the branches of dead trees, drilling a hole for each acorn (Figure 1.11). Where oak trees are abundant, acorn woodpeckers form very stable social groups, communally defending their larders, which provide group members with food during the winter.

The time for which food is stored varies from one species to another and depends, in part, on its perishability. Coal tits (*Parus ater*) store insects, having removed the head and the gut, in crevices in tree bark for a few hours. They also store pellets, each one consisting of 20–50 aphids compressed together. Owls store mice for a matter of days and jays and crows store seeds for a year. Ravens store meat for several weeks during the winter, burying it under snow, which keeps it frozen and so delays decay.

The energy that animals invest in food hoarding, and the importance of their cache for their winter survival is illustrated by the following observations:

- American biologist Bernd Heinrich observed an individual jay cache 127 food items in a single day.
- Individual Clark's nutcrackers (*Nucifraga columbina*), an American bird, collect up to 95 pine seeds at a time and hide them in locations as much as 22 km from their nest.
- Clark's nutcrackers breed in late winter, when there is no food available, and are totally reliant on cached food to rear their young.

Scatter hoarding presents animals with a problem: if they hide their caches from competitors, how do they find them themselves? Leaving clues, such as scent marks, would lead rivals to the hidden food, and so they must rely on memory. Thick-billed nutcrackers (*Nucifraga caryocatactes*) store 15–20 hazelnuts in each cache and are able to find them even after thick snow has covered them. Detailed observations of birds in winter suggest that some species are able to memorize the exact location of several thousand stored food items over several months. David Macdonald of Oxford University studied caching in the red fox (*Vulpes vulpes*) by rearing a pet vixen which he took for walks in the countryside on a long lead. She frequently cached food items that he gave her, burying them in the ground. Over the course of several weeks, she subsequently found 96% of them and only rarely did she dig in places where she had not hidden food.

How do we know that animals rely on what appear to be extraordinary feats of memory to find their food caches? An alternative explanation is that they

leave some kind of marker that neither their competitors nor humans can detect. Such an issue can only be resolved by carrying out experiments. Caching behaviour has been studied in American chickadees (*Parus* spp.) and European marsh tits (*P. palustris*). Captive-bred birds were trained to feed and hide food in an aviary containing artificial trees that had holes drilled in them, covered with Velcro lids. Opening these lids was very similar to the birds' natural behaviour of stripping bark off trees to find food. In the autumn, the birds spontaneously started to cache seeds in these holes. In a typical experiment, an individual bird was given 15 seeds to hide among a total of 72 open holes. When the bird had left the aviary, the seeds were removed and all the holes were covered up. Twenty four hours later, the bird was allowed back into the aviary. With very few errors, it visited the 15 holes where it had stored seeds the previous day, ignoring the others.

- Why did the experimenters remove the seeds before the bird was allowed back into the aviary?

- To eliminate any possibility that the bird was using any cues coming directly from the seeds, such as odour.

Scatter-hoarding requires a remarkable memory for the location of a large number of items, a process called spatial memory. Anatomical studies of the brains of birds have revealed that the part of the brain that is involved in spatial memory, called the hippocampus, is relatively larger in species for which scatter-hoarding is particularly important. For example, the hippocampus of the marsh tit (*Parus palustris*) is significantly larger than that of its close relative, the blue tit (*P. caeruleus*), which is not a scatter-hoarder.

### CHANGES IN SOCIAL BEHAVIOUR

For some animals living in temperate habitats, the onset of winter also brings about a marked change in their social behaviour. Most noticeably, many birds, having spent the spring and summer as breeding pairs, being highly aggressive towards all members of the same species, other than their mate, become gregarious and form large, cohesive flocks in which individuals behave cooperatively in a number of ways. As with other biological characters, sociality confers both benefits and costs on individuals.

A major benefit of flocking is that it reduces an individual's risk of being killed by a predator. There are three main reasons for this protection. Firstly, flocks detect predators more effectively than single birds. Secondly, predators become confused when attacking prey gathered closely together. Thirdly, members of flocks may defend themselves effectively against a predator even though, individually, they are not strong enough to do so.

Being in a social group increases feeding efficiency in other ways. Firstly, individuals gain information about the whereabouts of food by observing their fellow flock-members. Secondly, animals in groups can often disturb hidden animal prey more effectively than single individuals. Thirdly, predators in groups can hunt and kill prey that are too large for them to cope with on their own. Studies of captive great tits (*Parus major*), using an aviary similar to that described

earlier, showed that individual tits looking for hidden food found it more quickly when released into the aviary in a small flock. Individuals observed one another and avoided places where other birds had not found food, concentrating their search instead on places similar to those where other birds had located food.

For some animals, joining a social group in winter can reduce heat loss, especially on cold nights. For example, pallid bats (*Antrozous pallidus*) expend less energy during roosting when they roost huddled together than they do if they roost alone.

There are two major costs to becoming gregarious in winter. First, close association between individuals increases infection rates by parasites and pathogens. For example, in colonies of American prairie dogs (genus *Cynomys*), there is a positive correlation between the number of animals in a colony and the abundance of external parasites. Secondly, living in a group brings animals closer together, increasing the likelihood that they come into competition over food and other resources. This cost is borne particularly by smaller and weaker individuals, which generally lose in competitive interactions with other group members.

### SUMMARY OF SECTION 1.3

1 Coniferous trees are an example of plants that remain active during winter, with adaptations such as reduced water loss.

2 Endothermy allows some birds and mammals to remain active during cold winters, but places physiological and energetic demands on the organism, e.g. the need to maintain a high metabolic rate and/or reduce heat loss.

3 Bird plumage and mammal hair are highly effective insulators, reducing heat loss in winter. Heat loss is also controlled by heat exchange mechanisms and reduction in blood flow near the body surface.

4 The energy needed to sustain a high metabolic rate may be stored through either physiological and biochemical processes (adipose tissue) or changes in behaviour (hoarding or caching of food).

5 Some animals change their social behaviour during winter, becoming more gregarious.

## 1.4 STRATEGY 2: DORMANCY IN WINTER ('OPT OUT')

### 1.4.1 DECIDUOUS TREES

During the winter months, a combination of factors, including lower temperatures, reduced light intensity and shorter days, means that plants can only photosynthesize at a slow rate and for restricted periods. As a result, photosynthesis cannot produce energy as fast as respiration expends it. In addition, water is often in short supply because of freezing, and so plants that do not have adaptations to conserve water, as conifers do, would lose water. Deciduous trees avoid these problems in winter by dropping all their leaves and shutting off photosynthesis. Before they do so, they dismantle the photosynthetic apparatus in their leaves and withdraw many of the constituents to their branches, trunks and roots.

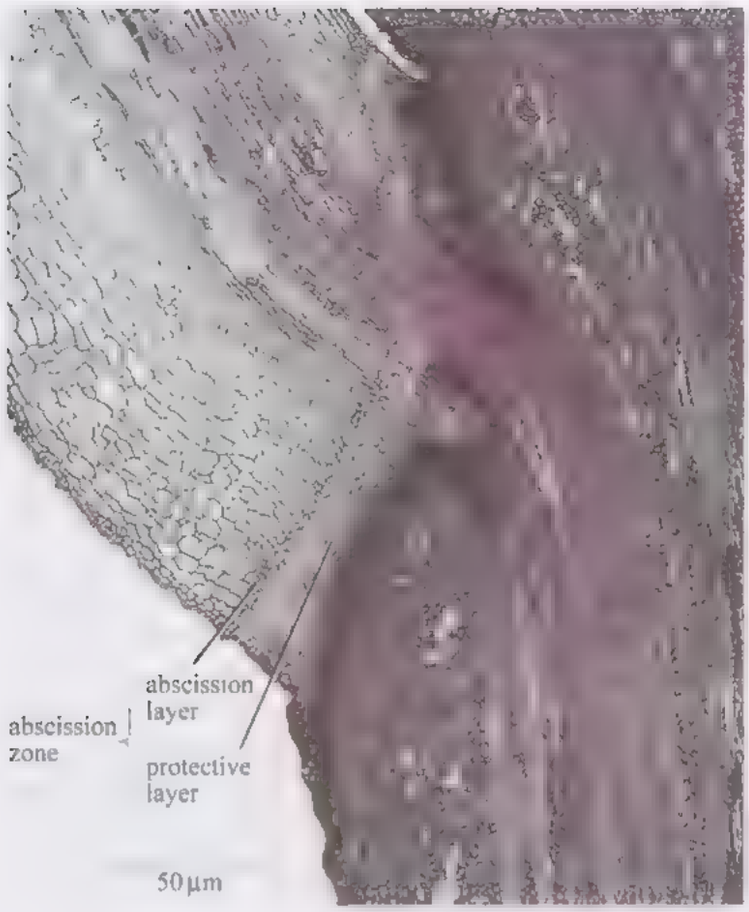



**Figure 1.12** A section through the base of the petiole of a maple (*Acer* sp.) leaf before abscission, viewed under the light microscope after staining. The abscission zone comprises a layer of cells where the leaf becomes detached and a protective layer which seals the exposed surface of the stem following abscission.

Thus, in the autumn, sugars, amino acids and such minerals as nitrogen, phosphorus and potassium are transported from the leaves to woody tissues. Chlorophyll is broken down and the products are also withdrawn from the leaves. It is this process that causes leaves to change colour in the autumn. The breakdown of chlorophyll leaves behind other pigments, such as orange carotenes and yellow xanthophylls, which are normally hidden by the green chlorophyll. Once as many nutrients as possible have been withdrawn from a leaf, an **abscission** zone forms where the leaf stalk (petiole) meets the stem (Figure 1.12). Here the vessels that supply water and nutrients to the leaf are closed off and the leaf detaches, leaving a protective covering of cork over the scar. Leaf abscission is controlled by a complex system of hormones, responding to lower temperatures and light intensity and to shorter day length.

### 1.4.2 WINTER STORAGE IN PLANTS

Many plants that survive winter in a dormant state form storage organs below the ground which store nutrients during the winter, the rest of the plant withering away. Storage organs come in a variety of forms, including tap roots, bulbs, corms, rhizomes, root tubers and stem tubers (Figure 1.13). In the carrot, the root is greatly enlarged into a fleshy tap root; the bulbs of onions are modified leaves; crocus corms, iris rhizomes and dahlia tubers are modified stems, and the tubers of potatoes are modified tips of underground stems.

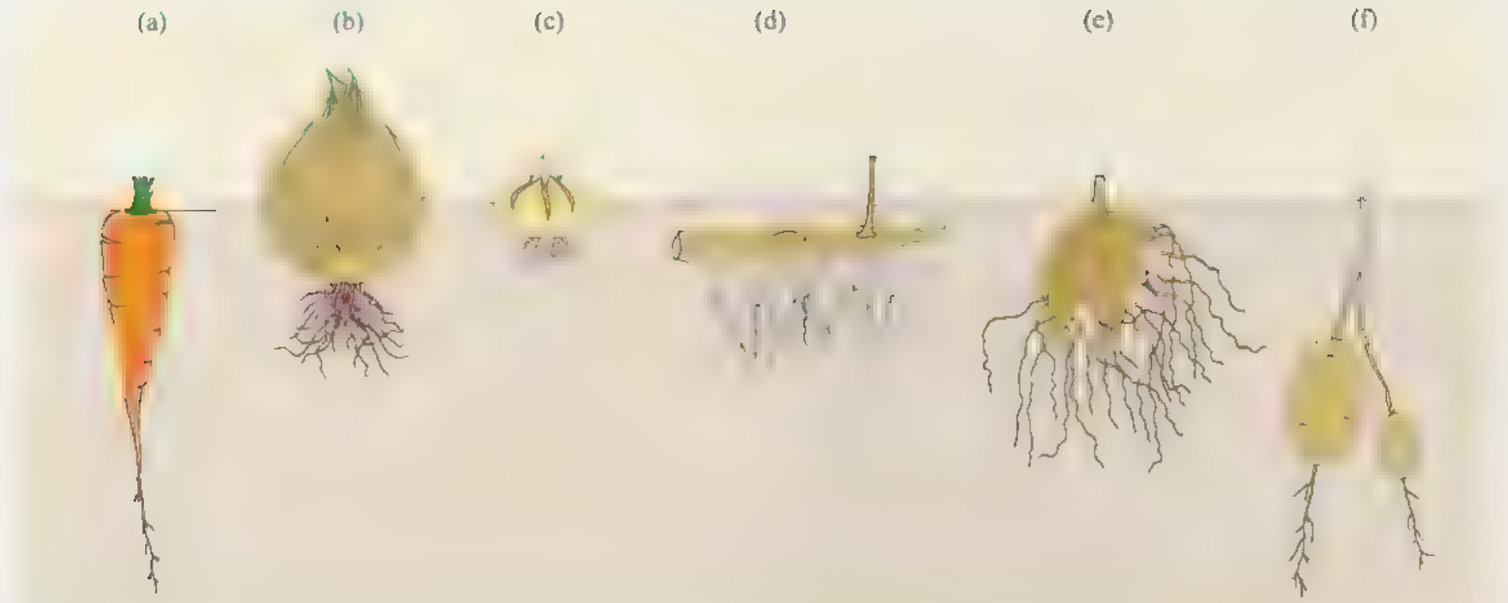



**Figure 1.13** Some examples of winter storage organs in plants: (a) tap root of carrot (*Daucus carota*, subsp. *sativus*); (b) bulb of onion (*Allium* sp.); (c) corm of crocus (*Crocus* sp.); (d) rhizome of iris (*Iris* sp.); (b) root tuber of dahlia (*Dahlia* sp.); (f) stem tuber of potato (*Solanum tuberosum*).

Unlike animals, few plants store energy reserves as fats (lipids). Those that do, generally store fats in seeds or fruits, a good example of the latter being the avocado (*Persea americana*). Storage in root organs is generally in the form of starch. Because they bind water, carbohydrates prevent desiccation of the storage organ and also reduce its freezing point, acting as 'antifreeze'. Plant storage organs provide ready-made larders for those herbivores that remain active in winter.

The root of a carrot serves as a storage organ, enabling the plant to complete its two-year life cycle. The storage organs of many plants are, however, also a means of asexual or vegetative reproduction. For example, the rhizomes of irises grow and branch and, as older parts of a rhizome die, two or more new plants are formed from the parts that are left. The tubers of a potato plant can each grow into a new plant and the bulbs of such plants as onions and daffodils (*Narcissus* sp.) divide to produce new bulbs and thus new plants.

### 1.4.3 Freeze tolerance in ectothermic vertebrates

In Britain, the vertebrate class Amphibia is represented by frogs, toads and newts. Amphibians are ectotherms, meaning that they are unable to generate large quantities of heat within their bodies, so their body temperature is close to that of their surroundings. The majority of amphibian species avoid the lethal consequences of being frozen, by digging their way under a large object, such as a rock, or deep into the soil, below the level that is penetrated by frost. There are some species, however, that have evolved a physiological response to very cold weather that enables them to survive the winter on or close to the ground surface. Examples include the American wood frog (*Rana sylvatica*) and the Asian salamander (*Hynobius kyserlingi*), both of which have distributions that extend far north of the Arctic circle. What they do is to infuse their tissues with antifreeze.

In the wood frog, the onset of cold causes the animal to become immobile. As the temperature falls below 0 °C, water in its toes begins to freeze. Within 10–15 minutes of freezing, glycogen stored in the liver is converted into soluble glucose which is released into the bloodstream, whence it finds its way into the cells and the extracellular spaces (Figure 1.14). The dissolved glucose lowers the freezing point of water, as antifreeze does in a car's radiator, preventing the formation of ice crystals and any consequent movement of water out of living cells.

Whereas the wood frog uses glucose as an antifreeze, the Asian salamander and the grey treefrog (*Hyla versicolor*) use glycerol, suggesting that this adaptation may have evolved independently in a number of amphibian species. Freeze tolerance allows these amphibians to survive freezing conditions for one or two weeks. It is not their only adaptation for surviving the winter: in the wood frog, for example, breathing ceases and the heart stops beating at very low temperatures.

Some ectothermic vertebrates rely on supercooling to survive short periods of cold temperature (see Section 1.2.1). For example, the spring lizard (*Sceloporus jarrovi*), living in the Arizona desert, survives very cold nights by supercooling. This strategy is risky, however, and many lizards die as a result of becoming frozen. Allowing its tissues to supercool is not a viable option for a frog or salamander; living in damp places, they are virtually certain to be in contact with ice crystals which act as nucleation points.

As well as enabling them to survive frosty conditions, the capacity to tolerate extreme cold confers other advantages on some amphibians. Many breed in temporary ponds that dry up early in the spring or summer, making it

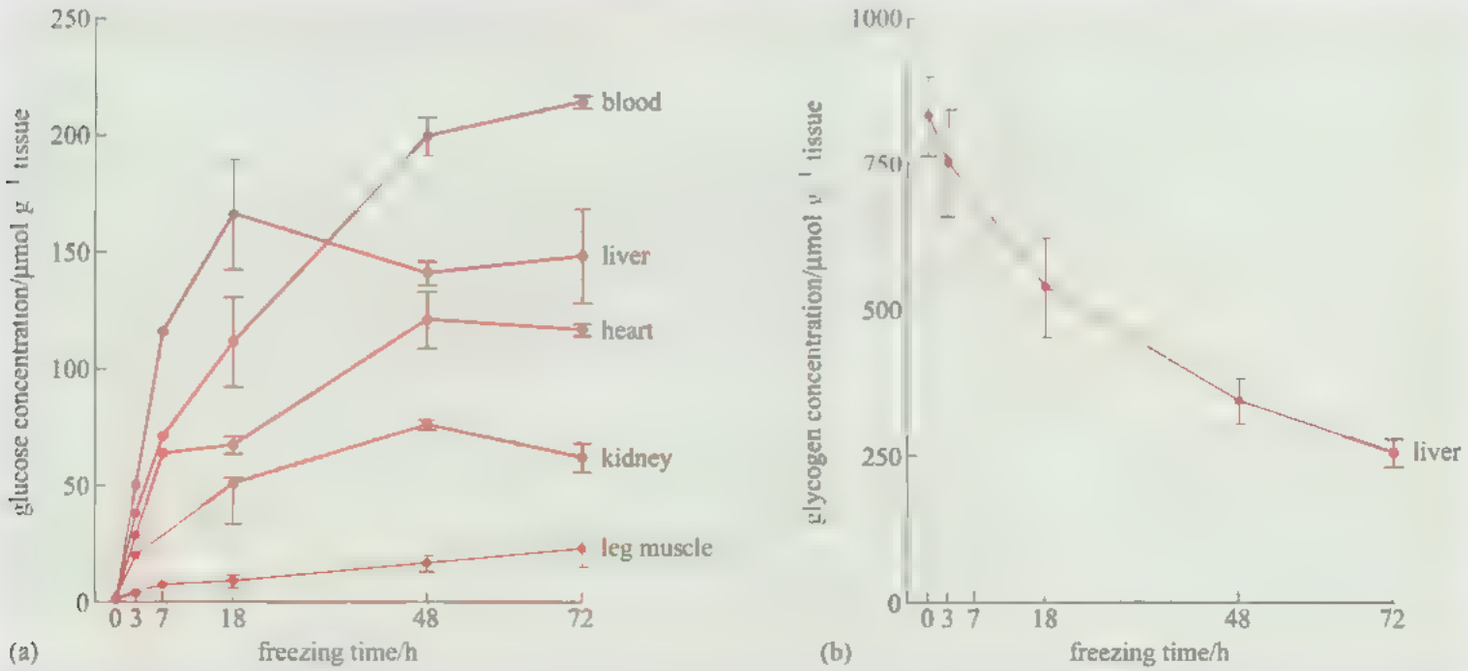



**Figure 1.14** (a) Changes in the concentration of glucose in various organs of the wood frog (*Rana sylvatica*) over 72 hours of freezing at −2.5 °C in the laboratory. (b) The corresponding depletion of liver glycogen reserves. Note that the horizontal axes are non-linear. Data from Pinder *et al.* (1992).

advantageous for breeding adults to migrate to ponds as early as possible in the spring. Early breeding maximizes the time available for the aquatic egg and larval stages to be completed before a pond dries up. Some species, such as the American blue-spotted salamander (*Ambystoma laterale*) migrate to breeding ponds while snow is still on the ground, giving them an advantage over other salamander species that do not start to breed until the weather is warm.

### 1.4.4 HIBERNATION IN MAMMALS

Many animals become inactive for periods of varying duration during the winter and there is a diversity of terms used to describe this state, including: sleep, torpor, dormancy, lethargy and hibernation. The word **hibernation** is often used loosely to refer to general inactivity but, in biology, it refers to a specific phenomenon, sometimes called 'true hibernation'. Hibernation is defined as the condition of passing the winter in a resting state of deep sleep, during which metabolic rate and body temperature drop considerably. It occurs only in certain mammals and one bird species, the poorwill (*Phalaenoptilus nuttallii*), a North American relative of the nightjar.

The phenomenon of hibernation is one reason why the term homeothermy is going out of fashion, to be replaced by endothermy, because maintaining a stable body temperature is the very opposite of what hibernators do. Instead, body temperature falls, from around 38 °C, to about 1 °C above ambient temperature, which is often close to 0 °C. At the same time, a hibernator's metabolic rate falls to as little as 1% of its normal value. The heartbeat becomes slow and irregular and breathing rate also slows.

Hibernation is an active process, that is, it is a state which animals enter into, not in response to immediate external conditions, but to internal stimuli. Some species are remarkably precise and predictable. For example, the arctic ground squirrel (*Spermophilus undulatus*) enters hibernation between 5 and 12 October and emerges between 20 and 22 April, regardless of the weather on those dates. This behaviour is in contrast to other winter states such as torpor or lethargy which are immediate responses to current conditions. Brown and black bears, for example, are lethargic during very cold periods but are otherwise active in the winter. A feature of hibernation that distinguishes it from other kinds of winter inactivity is that hibernators can arouse themselves spontaneously and are not dependent on external conditions, such as warm temperatures, to do so. The arctic ground squirrel is described as an **obligate hibernator** because it hibernates every winter. There are some mammals that are categorized as **facultative hibernators**, entering hibernation in response to very cold weather and poor food supply. The North American pocket mouse (*Perognathus californicus*) is a facultative hibernator.

True hibernation only occurs in relatively small mammals, though not all small mammals living in temperate habitats hibernate in winter, as we have seen. The largest mammal to hibernate is the marmot, which weighs about 5 kg. There are several reasons why larger mammals do not hibernate. Firstly, they would warm up too slowly and therefore use too much energy. Secondly, they have a smaller surface area to volume ratio and so can conserve body heat better than smaller

species. Finally, they are better able to carry a thick coat (Section 1.3.2) and sufficient adipose tissue to last through the winter. Hibernators are mainly found in the orders Rodentia, Chiroptera (bats) and Insectivora. The hedgehog (*Erinaceus europaeus*) is an example of a hibernating insectivore; in Britain, it hibernates from October/November to March/April. Note that although hedgehogs are in the order Insectivora they do not just eat insects!

The physiological features that are characteristic of hibernation are not maintained throughout the winter. Rather, the animal wakes up at intervals, its temperature and metabolic rate increasing to near-normal levels (Figure 1.15). The function of this periodic arousal is not wholly clear. Some species, such as the chipmunks (genus *Tamias*) eat from stored food reserves during arousal periods, but many others do not. Most species urinate and defaecate, move about and change their position, suggesting that arousal provides an opportunity for various essential physiological processes to be performed and to prevent the animal becoming moribund. From detailed measurements of Richardson's ground squirrel (*Spermophilus richardsonii*) in the laboratory, it has been calculated that, during the relatively brief periods of arousal (Figure 1.15), an individual expends 83% of all the energy that it uses up during the entire hibernation period.

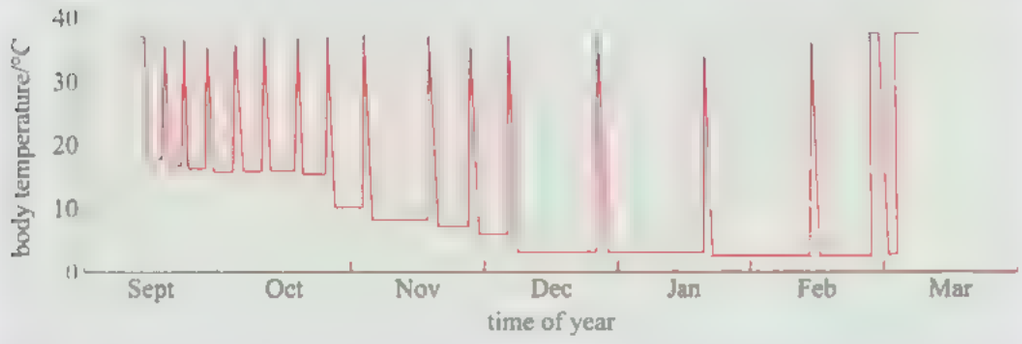



**Figure 1.15** Record of body temperature from September to March for a Richardson's ground squirrel. Data from Pough *et al.* (1996).

Hibernation requires internal energy reserves in the form of adipose tissue and hibernators typically feed intensively prior to winter, building up their fat stores. Some species, such as the edible dormouse (*Glis glis*), switch to a carbohydrate- and lipid-rich diet, e.g. seeds, at this time. A characteristic of hibernating mammals is that they possess larger quantities of a particular kind of adipose tissue called **brown adipose tissue (BAT)**. This tissue gets its name from its dark colour, which is due to the larger numbers of blood capillaries that permeates it and the high concentration of mitochondria within the cells. BAT is rich in mitochondria with special properties that enable it to oxidize fatty acids and/or glucose to produce heat very rapidly. BAT deposits are found around some internal organs and between the shoulder-blades of hibernators and their function is to generate body heat very rapidly, especially during periods of arousal.

Hibernation might seem to be a safe, and rather agreeable way to spend the winter but, for some species, it is fraught with danger. For Belding's ground squirrels (*Spermophilus beldingi*) living at high altitude in Tioga Pass, California, hibernation lasts 7–8 months. Two-thirds of all juveniles, hibernating for the first

time, and one-third of adult animals die during hibernation. Some die because their fat reserves run out before the end of hibernation; others are dug up and eaten by predators.

Some mammals spend the winter in groups, huddled together during periods of dormancy, and so conserve body heat. North American raccoons (*Procyon lotor*), for example, spend dormant periods in communal dens. Many species of bats hibernate communally. During hibernation, the body temperature of some bats can fall below 0 °C. In the autumn, they build up fat reserves that represent as much as a third of their total mass. During the winter, bats arouse themselves from hibernation to excrete and sometimes also to move to a new roost. A critical factor for hibernating bats is that roosting sites have high humidity and some populations have to migrate quite large distances to find suitable places, such as caves and hollow trees.

For some mammals, hibernation is closely associated with other important activities, notably reproduction and dispersal. Consequently, energy reserves may have to support more than one activity. For example, brown bears living at northern latitudes mate in the autumn and give birth to their cubs during winter lethargy. Edible dormice and some bats mate immediately after the end of hibernation. (In some species of bats, males wake up first and mate with the females before they have woken up!) The link between what animals do in winter and their reproductive cycles was discussed in Section 1.2.3.

**Natal dispersal** is the permanent departure of an individual from its place of birth, usually at the end of the breeding season. It is an important part of the life history of many animals, especially mammals, and tends to be sexually dimorphic, males dispersing further than females. Natal dispersal is potentially both hazardous and energetically expensive; dispersing animals tend to be vulnerable to predators and, being on the move, have little time to feed. Dispersal therefore requires internal energy reserves in the form of fat, the very same reserves that they later need to survive the winter. There may thus be a trade-off in the allocation of energy reserves to dispersal and to hibernation.

Scott Nunes of Michigan State University has studied dispersal and hibernation in Belding's ground squirrels in a locality where hibernation lasts for eight to nine months of the year. Young males typically disperse after the summer breeding season but show much variation in the extent to which they do so, with fatter males being more likely to move out of the natal area. In years when breeding is delayed, dispersal is inhibited; instead, young males remain near the natal area, building up their fat reserves prior to hibernation. The findings of this study are summarized in Figure 1.16.

If young males are ready to disperse early (date A in Figure 1.16), they leave with relatively small fat reserves. In this situation, they have time to travel further to a new area and then build up fat reserves before hibernation. If dispersal is delayed (dates B and C), males do not disperse unless they have built up a threshold level of fat reserves; the value of this threshold increases as winter approaches. In other words, the trade off between dispersal and hibernation is resolved by hibernation suppressing dispersal, unless an animal exceeds a certain fat level. After date C, dispersal is inhibited regardless of the size of the fat reserves.

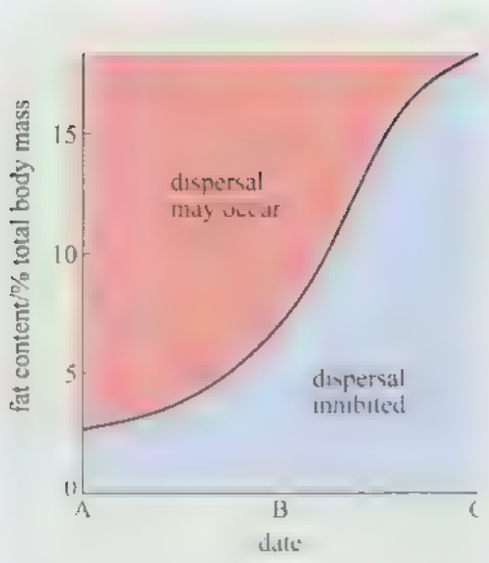



**Figure 1.16** A graphical summary of the control of dispersal in male Belding's ground squirrels. See text for explanation.

### SUMMARY OF SECTION 1.4

1 Deciduous trees avoid the problems of winter by shedding their leaves.

2 Plants can store nutrients over winter in a variety of structures.

3 Amphibians have evolved behavioural responses (e.g. burying themselves) and physiological responses (e.g. different types of antifreeze in the body fluids) to winter.

4 Hibernation occurs only in certain small mammal species and one species of bird and is accompanied by marked physiological and behavioural changes

5 Prior to hibernation, animals build up their fat reserves and frequently possess larger amounts of brown adipose tissue than non-hibernators.

6 There may be a trade-off between hibernation and dispersal in some animals.

## 1.5 STRATEGIES 3 AND 4: JUVENILE SURVIVAL AND MIGRATION

### 1.5.1 JUVENILE SURVIVAL

For organisms that are able to complete their life cycles within a year there is the possibility of spending the winter in various juvenile stages. We have already considered annual plants, the adults of which may die before the onset of winter, with seed not germinating until the spring. Surviving the winter as seeds has the advantages that the seeds are robust, and because they have a low water content they are less affected by freezing temperatures. Disadvantages of this strategy include the possibility that the seeds may be consumed by ground-feeding herbivores and that newly germinating plants may lose out in competition with annuals that germinated in the previous autumn.

Butterflies are good examples of insects that survive the winter in a variety of immature stages. Indeed, in Britain, different butterfly species overwinter at each of the juvenile stages of the life cycle, i.e. as eggs, larvae and pupae, or as adults.

Table 1.2 lists the numbers of species, categorized by butterfly family, that overwinter in each of these four stages in Britain.

**Table 1.2** The variety of overwintering strategies adopted by various juvenile stages (strategy 3) and adults (strategy 2) of butterflies in Britain. Data are numbers of species. Several species also migrate to Britain (strategy 4) and lay eggs. However, few individuals complete their development and return south; hence migration is not an important strategy for butterflies breeding in Britain.

| | Juvenile stage | | | Adult |
|---|---|---|---|---|
| Butterfly family | Egg | Larva | Pupa | |
| Satyridae | 0 | 11 | 0 | 0 |
| Nymphalidae | 1 | 9 | 0 | 4 |
| Lycaenidae | 6 | 7 | 2 | 0 |
| Pieridae | 0 | 0 | 5 | 1 |
| Hesperiidae | 2 | 5 | 1 | 0 |
| other | 0 | 0 | 2 | 0 |
| total | 9 | 32 | 10 | 5 |

› Based on Table 1.2, what is the predominant overwintering stage of butterflies?

● Larvae (caterpillars), occurring in 32 out of 56 species (57%).

There appear to be some phylogenetic patterns in overwintering strategy amongst the butterflies, i.e. some butterfly families in Britain are restricted to certain strategies.

› Which butterfly families are restricted to overwintering at one immature stage in Britain?

● The Satyridae (browns) only overwinter as larvae and the Pieridae (whites) overwinter as pupae.

In other families, such as the Lycaenidae (blues) and Hesperiidae (skippers), there is a mixture of overwintering strategies. However, these butterfly families are large and heterogeneous, and more careful inspection of subfamilies indicates phylogenetic patterns. For example, most hairstreaks (family Lycaenidae) overwinter as eggs. Across families, it seems there are common factors in the types of species that overwinter as a particular stage. For example, all of the species that overwinter as eggs have only one generation a year and are on the wing (and therefore breeding) later in the year, mainly July and August.

### 1.5.2 STRATEGY 4: MIGRATION ('GO AWAY')

About 40% of the bird species that breed in Britain do not spend the winter there but migrate south, some to southern Europe, others much further afield. The swallow (*Hirundo rustica*), for example, may migrate as far as the Cape of

southern Africa. From one perspective, migrants are European species that avoid the northern winter by migrating to a less severe environment. On the other hand, the swallow can also be regarded as an African bird that migrates to northern latitudes to breed. Why should an African bird migrate thousands of miles to breed, exposing itself to obvious risks? Swallows suffer 67% annual mortality, much of it occurring during migration. Long-distance migration also involves a huge energetic cost. The answer is that northern latitudes provide very good breeding habitats for birds. In late spring and summer these habitats support a rich food supply, in the form of seeds, fruits and insects, and there are relatively fewer herbivorous and insectivorous species living at northern latitudes, so that competition for food with which to rear their young is relatively low. Evidence for this conclusion comes from data on the clutch size of birds of the same or related species breeding at different latitudes. Temperate-breeding birds lay much larger clutches than similar species breeding at more southerly latitudes. This effect is also seen within species: for example, in the European robin (*Erithacus rubecula*), mean clutch size is 6.3 eggs in Scandinavia, 5.9 in central France, 4.9 in Spain, 4.2 in north Africa and 3.5 in the Canary Islands.

Birds are not the only animals that migrate. Among mammals, some populations of caribou or reindeer (*Rangifer tarandus*) make mass migrations of more than 1000 km, covering up to 150 km per day, in search of good grazing. Some insects also migrate over long distances. For example the monarch butterfly (*Danaus plexippus*) migrates from Canada and northern USA to Mexico, flying 120 km per day. Some 30% of the monarch population, however, does not migrate but hibernates during the winter.

Migration in birds requires considerable physiological preparation. Some time before they leave, migrants increase their feeding rate and lay down fat reserves, amounting to between 20 and 50% of total body mass, depending on the species. At the top of this range are the ruby-throated hummingbird (*Archilochus colubris*), which flies across the Gulf of Mexico, and the sedge warbler (*Acrocephalus schoenobaenus*), which migrates from Britain to west Africa. Both these species complete their migration in a single sustained flight. Species that carry smaller loads of fat typically fly in a series of stages, stopping to feed and put on weight along the route. The rate at which migratory birds put on weight prior to migration is remarkable. Whitethroats (*Sylvia communis*) preparing to migrate from Sweden increase their food intake by about 70%, their body mass increases by 7% per day, and they depart 50% heavier than normal.

Prior to migration, smaller birds build up larger fat reserves, relative to their body size, than do larger birds. This size effect is related to a measure of the amount of power that a bird can generate, called the **power margin**, which is defined as the difference between the maximum power than can be developed by the flight muscles and the power required for unladen (with no extra fat) level flight at a standard speed in still air. A large bird such as a swan has a small power margin, so swans need to build up speed before they can take off, and, once airborne, they climb rather slowly. Small birds, with a large power margin, can take off vertically and climb very rapidly (Hedenstrom and Alerstam, 1992). The small power margin of large migrants, such as white storks (*Ciconia ciconia*) which

migrate from Europe to Africa and back, means that they cannot carry large amounts of fat to support long flights. They economize on fuel, to some extent, by soaring on thermals, rather than using flapping flight, but these habits constrain them to routes that do not cross the sea. Storks migrate either over Gibraltar in the west, or over Israel in the east, depending on where they breed. They also make frequent stops to feed.

As birds put on weight prior to migration, they start to show **migratory restlessness**. Birds held in captivity just before the migration season do a lot of hopping about and experiments have shown that this hopping is not random, but is oriented in different directions in spring and autumn (Figure 1.17).

Experimental studies of captive birds have revealed that the proximate factors that stimulate increased feeding prior to migration and migratory restlessness are decreases in the L : D ratio (i.e. longer nights) and, in some species, a circannual clock.

To complete their life cycle, migratory birds make at least two journeys: outwards to the wintering site and a return journey next spring back to the breeding site. A major determinant of reproductive success for many temperate bird species is the

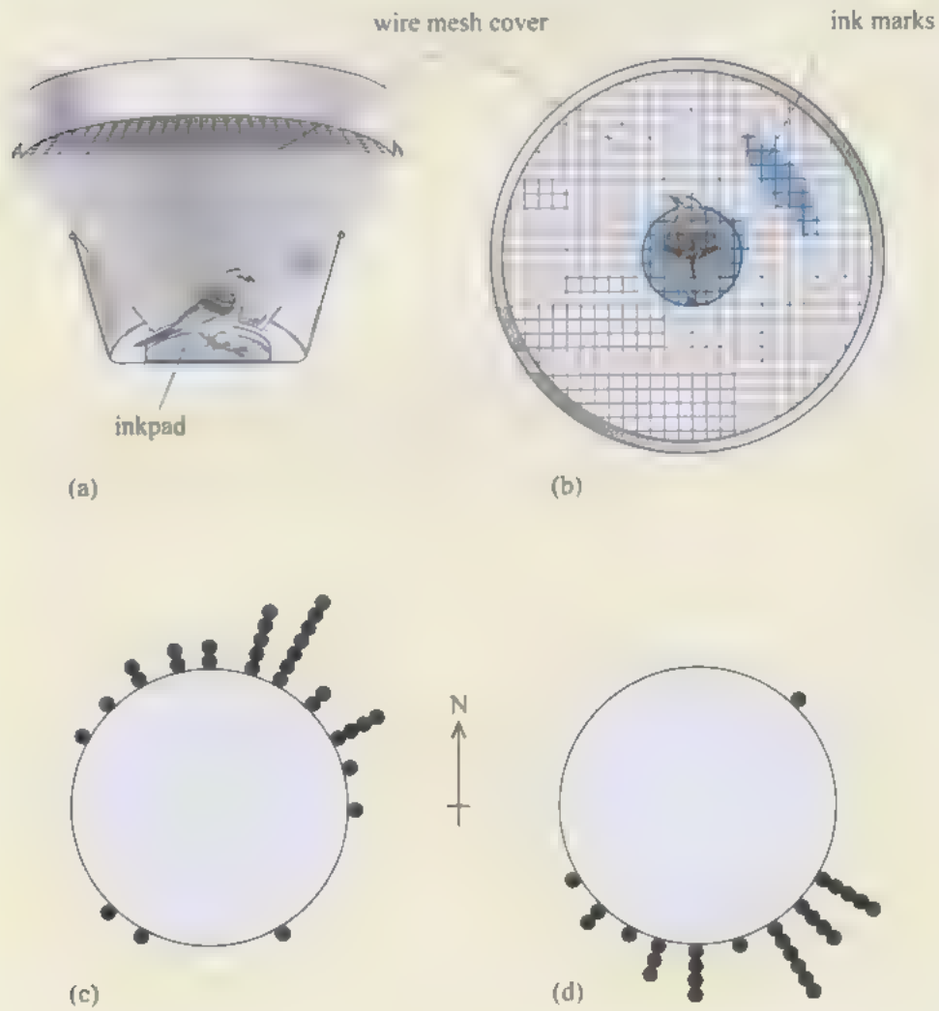



**Figure 1.17** Orientation of migratory restlessness in a caged bird. (a) The cage has an inkpad at the centre and a wire mesh lid. (b) View downwards into the cage. Every time the bird jumps from the inkpad towards the edge, its feet leave a mark. These marks are oriented in a northerly direction in spring (c) and towards the south in autumn (d). Data from Pough *et al.* (1996).

date on which they start to breed. In the majority of species, pairs that start to breed early in the spring fledge more young. There are several reasons for this correlation, the relative importance of which varies from species to species. Firstly, early breeders generally secure better territories, in terms of the abundance of food that they contain. Secondly, an early start often means that a pair has time to produce and rear a second clutch in the same season. Thirdly, in some species the earlier breeders are better synchronized with their environment, such that the time of peak food requirement (i.e. during feeding of the young) is synchronized with the time of maximum food availability.

A recent study of the American redstart (*Setophaga ruticilla*), suggests that a further factor could be important. It was found that the earliest birds to arrive at breeding sites were in better condition than later birds, some of which arrived up to a month later. This study investigated the tropical wintering habitat of this species and found that the early-arriving, good-condition birds had wintered in better habitats than later arrivers. This study therefore suggests that the quality of winter habitats may be an important determinant of fitness for migratory birds. It also reinforces the point that, in considering the success of organisms, it is important to consider all parts of their life cycle.

### SUMMARY OF SECTION 1.5

1 Some annual plants and insects can spend the winter at juvenile stages, such as seed, egg, larva or pupa. Butterflies in Britain display a variety of juvenile overwintering strategies.

2 Migration often results in high mortality, but completion of the journey results in higher breeding success, due to increased availability of food and fewer competitors.

3 Birds increase their body mass, sometimes by up to 50%, prior to migration. The body mass increase is related to the power margin of the birds.

4 Early arrival of migratory birds at the breeding site confers important advantages on those individuals.

## 1.6 CONCLUSION AND OVERVIEW OF BOOK THEMES

This chapter has presented an overview of the ways in which organisms living in temperate habitats are adapted to survive the winter. The chapter has shown how a limited set of environmental changes associated with the onset of winter can lead to a diversity of adaptations and therefore a large diversity of species.

On the basis of the examples discussed in this chapter, we can identify four factors that contribute to the diversity of adaptive strategies for coping with winter.

1 *Alternative strategies and trade-offs* For any one environmental 'problem', there may be two or more solutions that may be appropriate in certain species under certain conditions. Thus, remaining active, becoming dormant and migrating are all viable options for long-lived organisms to cope with winter. Among animals that become dormant, building up reserves in the form of a larder is an alternative to depositing lipids in adipose tissue. These adaptive

solutions may be thought of as being traded off, i.e. natural selection may lead to an organism adopting one particular adaptive solution. This effect is reinforced by an organism's evolutionary history (see below).

2 *Phylogeny* The evolutionary history of different groups of organisms has equipped them with a range of adaptations that predispose them towards one strategy rather than another. For example, endothermy and thick fur make hibernation a viable strategy for mammals, wings make long-distance migration a viable strategy for birds.

3 *Size* Body size can have a marked effect on what organisms do in winter. Dying back and overwintering as an underground storage organ is a viable option for small plants, but not for large woody ones; hibernation is possible for small to medium-sized mammals but not for large ones.

4 *Life history* The particular life history of organisms allows them to survive the winter in different ways. Thus annual plants have the option of overwintering as small plants or seed whilst herbaceous perennial plants have the extra option of surviving the winter underground. As life history is itself subject to natural selection, it is another example of evolutionary history restricting the options of adaptive solutions.

A fifth factor, which we have not had the space to explore in this chapter, is coevolution. This concept will be developed in Chapters 4 and 5. Through coevolution, adaptations by one group of organisms create opportunities for other groups. Thus, in the context of the present chapter, if it were not for the fact that many plants store food over the winter, there would be little opportunity for many herbivores to remain active in winter. Coevolution implies that, if selection favours herbivores that can exploit underground storage organs, then, equally, selection also favours plants that are able to defend these resources.

These five factors serve as major themes for this book in which we attempt to understand the *processes that generate diversity*. Chapter 2 concerns ways that heterotrophic organisms acquire and process food. Food is important, as the supply of energy and nutrition is the basis upon which activities such as growth and reproduction can be traded off with other biological processes, such as defence. In Chapter 3 we review the genetic basis of the diversity of organisms. This chapter is a key one, as all adaptation and variation, and therefore all biological diversity, has a genetic basis. Chapters 4, 5 and 6 cover three major processes in the life history of organisms: reproduction, defence and longevity. All three processes are traded off within an organism's life history. Thus for a given energy input, a high energy burst of reproduction may be associated with subsequent low levels of survival, as discussed in Chapter 6. These trade-offs within an organism's life history are reflected in the range of strategies evolved by individuals of different species. Chapter 4 shows how the bewildering array of reproductive strategies has been generated and, in turn affects the generation of genetic diversity described in Chapter 3. In Chapter 5 we see strong evidence for coevolution generating diversity in hosts and their pathogens.

## REFERENCES

Baker, J. M. R. (1992) Body condition and tail height in great crested newts, *Triturus cristatus*, *Animal Behaviour*, **43**, pp. 157–159.

Bronson, F. H. (1987) Environmental regulation of reproduction in rodents. In: *Psychobiology of Reproductive Behavior*, D. Crews (ed.), Prentice Hall, New Jersey. p. 209.

Hedenstrom, A. and Alerstam, T. (1992) Climbing performance of migratory birds as a basis for estimating limits for fuel-carrying capacity and muscle work, *Journal of Experimental Biology*, **164**, pp. 19–38.

Irving, L. (1966) Adaptations to cold, *Scientific American*, **214**, pp. 94–101.

Kenagy, G. J. (1981) Effects of day length, temperature, and endogenous control on annual rhythms of reproduction and hibernation in chipmunks (*Eutamias* spp.). *Journal of Comparative Physiology A*, **141**, pp. 369–378.

Macdonald, D. W. (1976) Food caching by red foxes and some other carnivores, *Zeitschrift fur Tierpsychologie*, **42**, pp. 170–185.

Pinder, A. W., Storey, K. B. and Ultsch, G. R. (1992) Estivation and hibernation, in: *Environmental Physiology of Amphibians*, Feder, M. E. and Burggren, W. W. (eds), Chicago University Press, pp. 250–276.

Pough, F. H., Heiser, J. B. and McFarland, W. N. (1996) *Vertebrate Life* (4th. edn), Prentice Hall, New Jersey, pp. 589, 688, 695.

Nunes, S., Ha, C.-D. T., Garrett, P. J., Mueke, E. M., Smale, L. and Holekamp, K.E. (1998) Body fat and time of year interact to mediate dispersal behaviour in ground squirrels, *Animal Behaviour*, **55**, pp. 605–614.

Valerio, P. F., Kao, M. H. and Fletcher, G. L. (1992) Fish skin: an effective barrier to ice crystal propagation, *Journal of Experimental Biology*, **164**, pp. 135–151.

## FURTHER READING

Barnes, R. S. K., Calow, P. and Olive, P. J. W. (1988) *The Invertebrates: a New Synthesis*, Blackwell Scientific Publications, Oxford. [A comprehensive review of diversity, evolution and ecology of invertebrates.]

Irving, L. (1966) Adaptations to cold, *Scientific American*, **214**, pp. 94–101. [An important and accessible summary of physiological adaptations of vertebrates to cold.]

Krebs, J. R. and Davies, N. B. (1987) *An Introduction to Behavioural Ecology* (2nd edn), Blackwell Scientific Publications, Oxford. [A collection of essays by leading researchers on key themes in behavioural ecology.]

Pond, C. M. (2000) *The Fats of Life*, (2nd edn) Cambridge University Press. [This book explains in simple language the biology of feasting and fasting, fattening and slimming in wild animals as well as people.]

Pough, F. H., Janis, C. M. and Heiser J. B. (1999) *Vertebrate Life* (5th. edn), Prentice Hall, New Jersey. (6th edn expected 2001) [A wide-ranging review of the diversity, function and evolution of vertebrates.]

Raven, P. H., Evert, R. F. and Eichhorn, S. E. (1999) *Biology of Plants* (6th edn.), Freeman, New York. [A richly-illustrated book that deals with all aspects of plant biology, from cellular mechanisms to ecology, and covering physiology, evolution and diversity.]

# DEALING WITH FOOD

## 2.1 INTRODUCTION

Organisms need materials with which to build the fabric of their bodies and fuel to power movement and the synthesis of secretions, growth and repair of tissues, and the production of gametes for the next generation. These interconnected processes are collectively known as metabolism, and they depend upon adequate supplies of carbohydrates, lipids and proteins of appropriate composition, and smaller quantities of certain metallic ions, notably sodium, potassium, magnesium and calcium, and organic molecules which, when derived from the diet, are known as **vitamins**. The principles of metabolism are broadly similar in all animals, but how much material can pass through any particular pathway, and the rates of the processes, differ enormously between species.

Heterotrophs obtain all their energy supplies and most of their other nutrients from the tissues of other organisms. Animals and fungi are the main multicellular heterotrophs, though there are a few lineages of carnivorous plants. Fungi more often feed off dead or dying organisms, and animals actively kill living plants and animals, but there are many exceptions. Carnivorous plants fall outside the definition of typical heterotrophs because they make organic molecules from inorganic carbon dioxide by photosynthesis, like typical green plants. But they resemble animals in obtaining nitrogenous materials by catching and digesting insects and other small invertebrates. They therefore share many of the problems faced by carnivorous animals: how to obtain nutrients from organisms that actively avoid their predators.

The key features of heterotrophy are the production of digestive enzymes which break down complex organic materials, absorption of the breakdown products, and getting near enough to the food source for these processes to be efficient. Heterotrophy is not an 'easy option' compared to autotrophy: obtaining and processing other organisms as food are energetically expensive activities requiring elaborate anatomical and biochemical adaptations. For most animals, many of their tissues, and a large fraction of the total energy they expend, are devoted to foraging and to the digestion and absorption of nutrients. With a few exceptions (fruit, nectar, etc.), organisms actively avoid making their tissues or secretions available as food for other organisms, at least while they are alive. They disperse, hide, run away, protect themselves with shells, spines or tough skeletons, or make themselves poisonous or distasteful. Some of these topics are discussed in Chapter 5.

The basic mechanisms of digestion and absorption are similar in all animals, though details such as the relative abundance of enzymes and the conditions under which they function most efficiently differ between species. In contrast, there are many different sources of food, and many ways of obtaining and processing it. The huge variety of animals, at least one million described species, is due mainly to the diversity of strategies for counteracting the food organism's defences against its eater, i.e. differences in diet, and in ways of finding, catching,

fragmenting and digesting food, and of avoiding being eaten themselves. This chapter outlines these fundamental and contrasting aspects of obtaining and dealing with food. The range of examples shows how the variety of solutions to the problems of nutrition and digestion contribute greatly to the diversity of organisms.

### 2.1.1 STRUCTURE OF THE MOUTH AND GUT

The gut is one of the defining characters of animals. It is a hollow sac or tube lined, except near its openings, with a layer of tightly linked cells called intestinal **epithelium** which both produces digestive enzymes and regulates nutrient uptake (Figure 2.1). The interior of the gut is thus morphologically part of the outside world, and its lining may be compared to a skin, controlling interactions between the contents and the body. The structure and anatomical relations of the gut are, together with limbs and other locomotory apparatus, among the most important features that distinguish animal phyla from one another. The relative masses of the guts and their component parts differ enormously between species, depending upon diet and digestive mechanisms.

In simpler invertebrates, ciliated cells, aided by wriggling movements of the whole body, stir the gut contents, bringing them into close contact with the lining and propelling them towards the anus, where undigested material is expelled. In vertebrates and some more complex invertebrates, including arthropods, layers of muscle (Figurer 2.1a) surrounding the epithelium perform these functions with waves of squeezing movements called **peristalsis**. These muscles may be very powerful, especially in animals that suck, or whose food consists of large, tough particles. For example, those of the South American bedbug *Rhodnius* generate forces of 3–6 times atmospheric pressure, sucking in mammalian blood (they prefer that of humans) at 4 m s$^{-1}$ through their syringe-like mouthparts, whose internal diameter is only 10 μm. Such power enables the insects to take in several times their own mass of blood in just a few minutes, thus minimizing the time for which they have to linger where their victim might swat them.

All vertebrates except the class Agnatha have jaws and most have teeth, enabling them to bite, tear, grind or chew large food items into small fragments. (In birds, tortoises and a few other groups, the teeth are replaced by a horny beak.) The structure of the teeth, beaks, mouthparts, tentacles, etc. and the muscles that operate them are intricately adapted to the diet, especially in animals that feed only on one or a few kinds of prey or food plant. The structure and operation of the mouthparts, beaks or teeth often provide the best clues about an animal's diet and food gathering strategy. In many vertebrates, the jaw muscles are among the most powerful and accurately controlled muscles in the body, and many birds, including parrots, woodpeckers and weaver-birds, as well as certain mammals, use the mouth for tasks such as nest-building, which involve strength, speed and skill, in addition to food manipulation.

Arthropods do not have true jaws like vertebrates, but they have one or more pairs of limbs called **mouthparts** in front of, behind or beside the mouth, which are specialized for food manipulation. They are enormously diverse in structure, adapted to grasp, tear, filter, wipe, bite, chew, rasp, suck and inject poison.

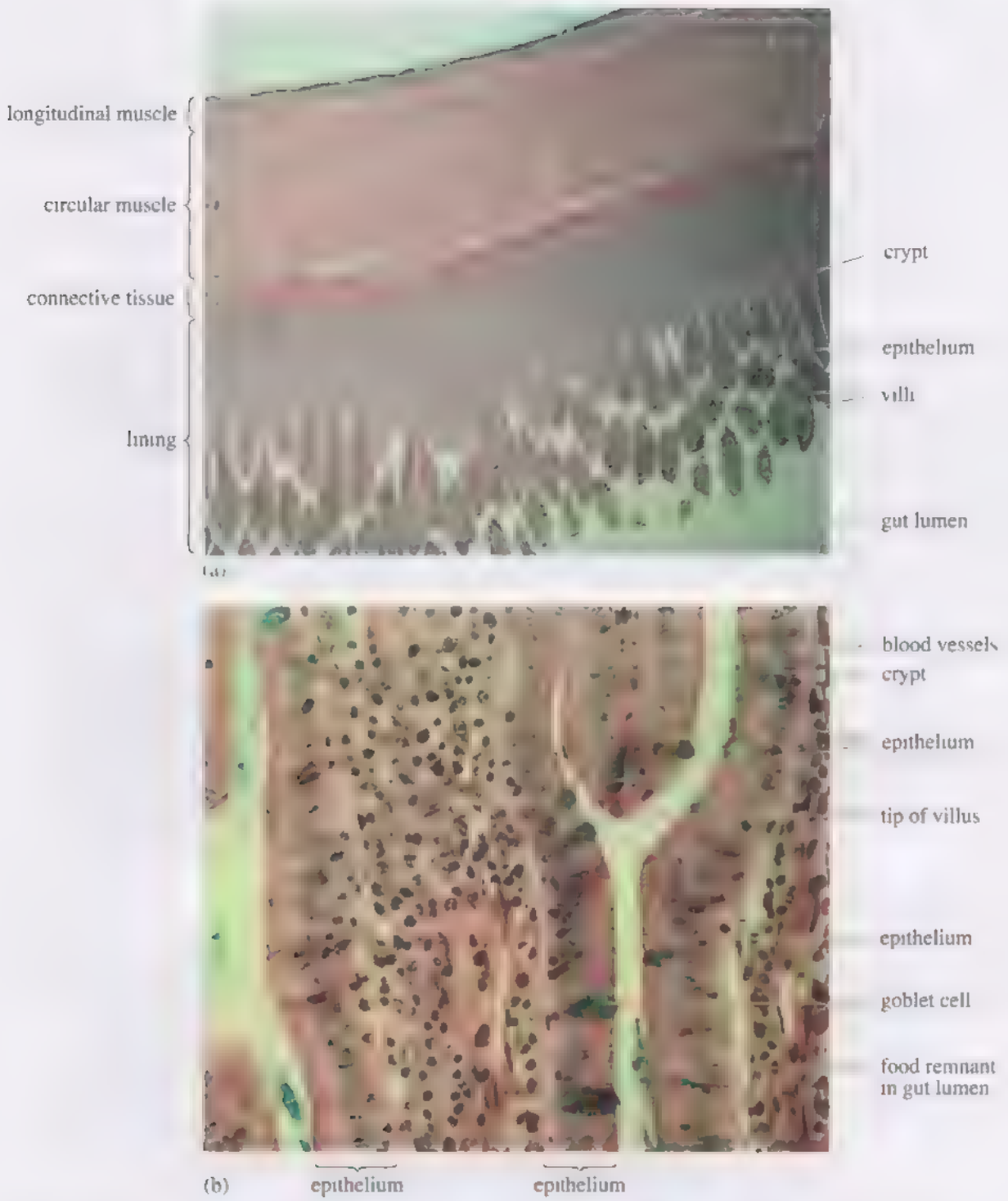
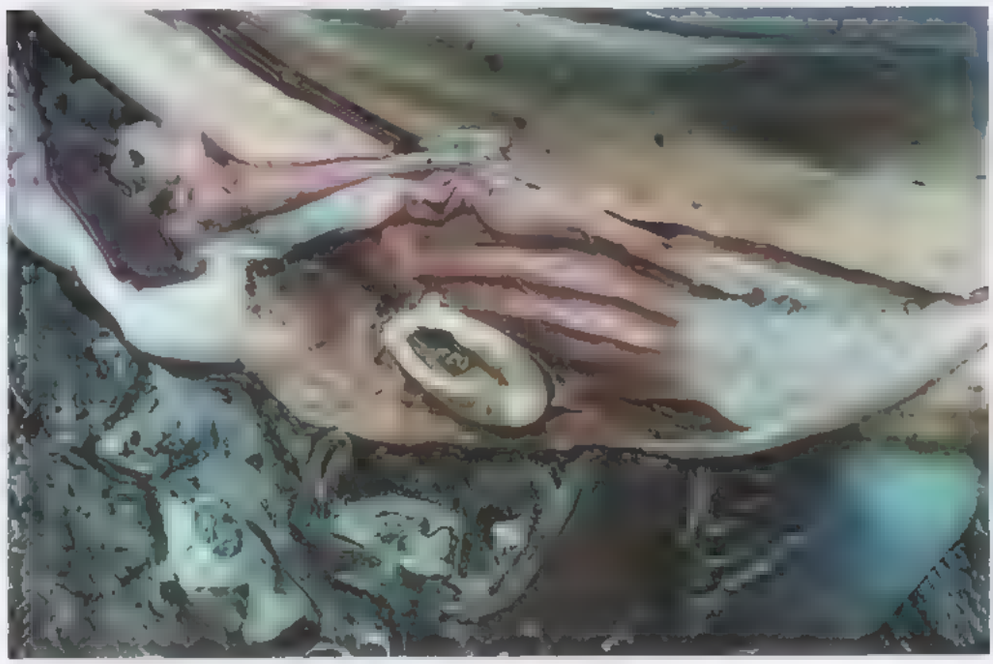
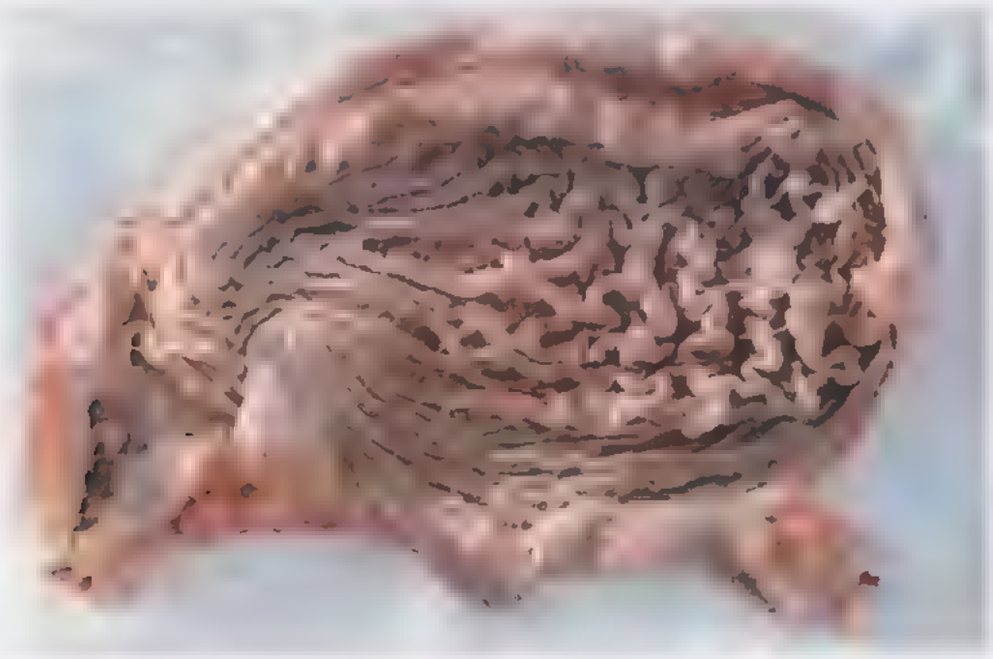



**Figure 2.1** Cross section of the small intestine of a rat. (a) Magnified 25 times. The two layers of powerful muscles are clearly visible, and many of the finger-like projections (villi) in the lining have been sectioned at a variety of angles. (b) When magnified 400 times, the continuous layer of epithelium is clearly seen, containing dark stained goblet cells, which secrete mucus, and other cells that secrete digestive enzymes. The core of each villus contains numerous blood vessels. The green-staining fragments between the villi are the remains of the rat's final meal.

Mechanical fragmentation of food increases the surface area on which chemical breakdown of large molecules can take place, enabling digestion to be faster and more complete. It begins in front of or inside the mouth, where **saliva**, a watery fluid containing mucus to make particles slippery, and sometimes enzymes, is secreted as soon as eating begins (or even in anticipation of it!). The food passes through the **oesophagus** to the **stomach**, where muscular churning movements continue its fragmentation, and digestive enzymes and, in vertebrates, acid are secreted onto it. The lining of the stomach is often pleated (Figure 2.2), which maximizes the area of secretory epithelia, with regions of contrasting appearance producing different secretions.



(a)



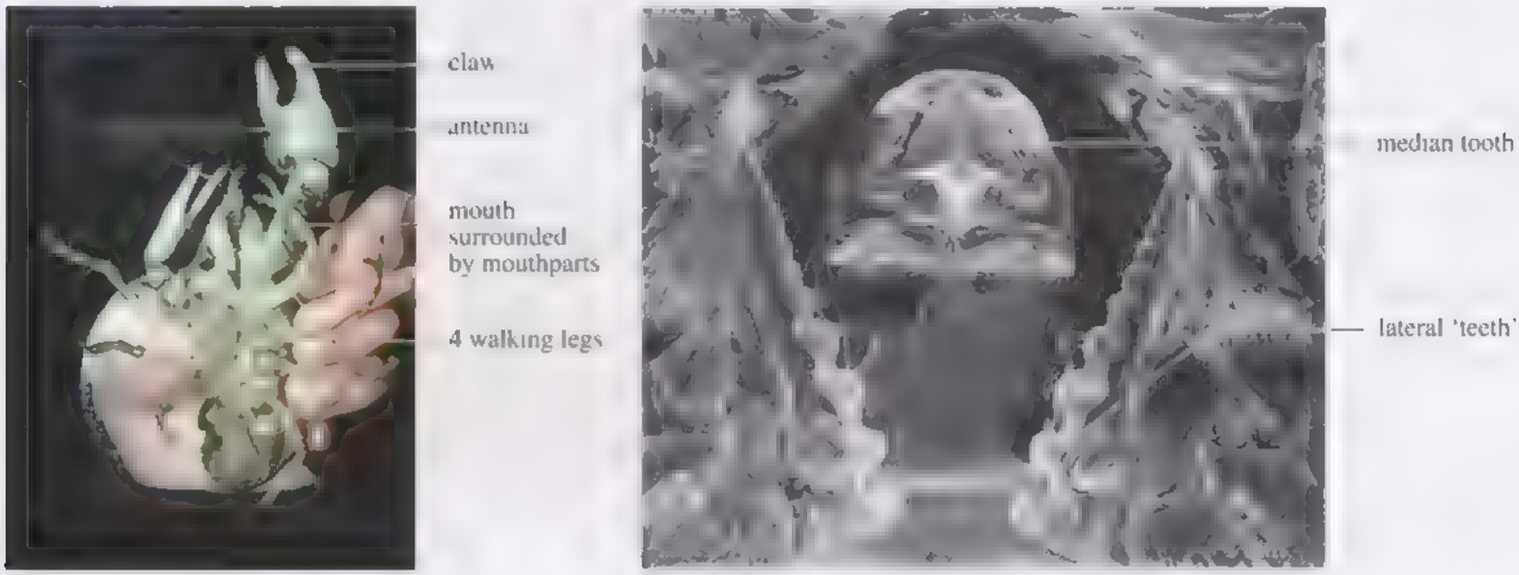
(b)

**Figure 2.2** The inside of the stomach of (a) a dogfish (*Squalus* sp.) and (b) a polar bear (*Ursus maritimus*). In (a) note the distinct appearance of different regions of the stomach and its pleated surface, which increases the area of contact with the food and allows for expansion. (b) This animal's stomach has shrunk (to about 50 cm across) because it has been fasting for several weeks, but the flexible walls enable it to expand enormously.

) In which terrestrial vertebrates would breaking up food in the stomach be important?

● Birds. They do not have teeth, although many are predators and/or eat hard seeds.

Part of the avian stomach forms a massively muscular **gizzard** which grinds tough food items, sometimes with the help of hard grit or pebbles which the bird swallows deliberately. The stomach has a similar role in many large invertebrates. Crabs and crayfish shred their food with their large claws and several pairs of stout mouthparts, ingesting the fragments one at a time through their relatively small mouth (Figure 2.3a). Mechanical breakdown continues in the stomach (called the proventriculus) which is equipped with hard 'teeth' and powerful muscles, shown in Figure 2.3b.




Beyond the stomach is the **intestine**, which is devoted to chemical digestion of food and the absorption of the small molecules so released. The digestive enzymes and other substances secreted by the gut epithelium are supplemented in many larger, more complex animals by those produced in associated glands, including the liver and pancreas in vertebrates. In simple animals, the whole gut has a similar structure but the presence of these additional glands, and specialization of the intestine itself, define distinct regions in vertebrates and large arthropods. The 'small' intestine is narrow but long and often coiled or folded. The 'large' intestine, also called the hindgut or the **colon**, is usually much shorter but wider and is specialized for absorbing salts, and with them water, from the gut contents. Maximizing the area of epithelium in contact with the food makes digestion and absorption faster and more efficient, so the lining of vertebrate guts has finger-like processes called **villi** (singular villus) which project into the gut contents (Figure 2.1), especially in absorptive regions. Any material remaining after digestion and absorption are complete is expelled through the anus as **faeces**.

**Figure 2.3** (a) Ventral view of an adult crayfish, *Austropotamobius pallipes*. This native species used to be common in streams and shallow rivers in southern England, where it eats soft vegetation and picks flesh from dead animals. (b) Scanning electron micrograph of proventricular teeth in the lining of the stomach of a crayfish. The teeth are moved by the indirect action of strong muscles attached to the wall of the stomach, resulting in grinding of food between the lateral teeth and the median tooth

Although the overall movement of the gut contents is from mouth to anus, the rate of passage is highly variable, and reflux motion is common, especially in birds and herbivorous mammals. In vertebrates, the muscles associated with the gut differ from those of the limbs and trunk in their internal structure and in the physiological mechanisms that control them. The movements and secretions of the gut and its associated glands are controlled by neurons located in small groups (known as ganglia) that are in or near these organs. Those that lie within the gut wall are known as *enteric* neurons and form part of the *autonomic* nervous system, so their activities are outside voluntary control.

) How does this difference in innervation reveal itself in our ability to control and perceive these regions of the gut?

● We can control the mouth and anus because their nerves and muscles are similar to those of limbs, but we are largely unaware of and unable to control most of the gut.

Microbes are present in the guts of all free-living animals. Most are harmless, indispensable symbionts that aid digestion, especially by breaking down cellulose and toxic substances from plants. A few pathogens inevitably slip in with food, and the gut lining of most higher animals contains components of the immune system (see Chapter 5) which protect the delicate, permeable surface from disease.

### SUMMARY OF SECTION 2.1

1 Intestinal epithelium lines the gut except around the mouth and anus, and in vertebrates and advanced invertebrates is surrounded by muscles that propel and/or grind the food. The muscles are controlled by the autonomic nervous system.

2 Most vertebrates have jaws, and arthropods mouth parts, often armed with hard and/or sharp teeth, which fragment food before it is taken in. Gizzards and stomachs also grind large particles.

3 The stomach and intestine differ in muscularity, and secretory or absorptive capacity. In more complex animals, the secretory functions of the gut itself are supplemented by associated glands. The effective surface area of absorptive regions is greatly increased by numerous villi.

## 2.2 CARNIVORY

The majority of animal species are partly or entirely carnivorous (flesh-eating) but predators kill as well as eat other animals. Most animal tissues are nutritious and digestible to other animals, so the main problems for predators are catching and subduing their prey, rather than with digestion itself. As well as running away or hiding, potential prey protect themselves with shells and spines, and/or produce poisonous secretions which could injure the predators, and in some, the flesh itself is toxic. There is an almost infinite variety of styles of predation; as well as chasing or ambushing the prey, predators use traps, nets or lures, often combined with elaborate camouflage. Diversity of predation strategies is a positive advantage to the predators, as prey cannot hide from all possible predators all of the time.

Most fish and amphibians generally gulp small prey with little or no fragmentation, or tear it into chunks small enough to be swallowed. Large items may be tossed around until they are in a favourable orientation for swallowing. This style of feeding is retained in many carnivorous reptiles and birds, the main function of their teeth or beak is to pierce and hold the prey, rather than to chew or slice it.

Crocodiles and alligators belong to an ancient lineage of reptiles whose style of killing and dismembering large prey typifies that of predatory vertebrates. Mature crocodiles conceal themselves under water as they approach a mammal or large bird while it is drinking or swimming across a river, then lunge at it, usually from one side. The jaws and teeth are rarely strong enough to kill the prey by destroying a vital organ such as the brain or heart. Instead, the reptiles exploit the fact that they need to breathe air much less frequently than warm-blooded terrestrial animals: they drag the prey under water and hold it there until it drowns. The rigid skull and jaws (Figure 2.4) support stout, cone-shaped teeth and powerful jaw-closing muscles, enabling the predator to maintain a firm grip for several minutes. Struggling prey may cause injuries, but fish, amphibians and reptiles can replace broken teeth and add new ones throughout life.


**Figure 2.4** The anterior half of a crocodile skeleton. Note the long stout jaws and numerous simple, pointed teeth.

Once the prey is dead, the crocodile clamps its teeth around the edge of a limb or the neck, and flails its tail and legs, often turning itself right over in the process, until it rips off a morsel small enough to be swallowed. Several large crocodiles tearing a zebra or wildebeest to bits in this way make an impressive sight: almost every muscle in the crocodiles' bodies contributes to food collection. Fragmenting the prey is wasteful, of course, as blood and gut contents usually flow away, and only the smaller bones can be swallowed whole. Even with such vigorous pre-shredding of the food, digesting a large catch can take weeks. Juvenile crocodiles and alligators, which eat insects and small fish, feed frequently throughout the year, but large adults, which live mainly on mammals, may kill as rarely as once or twice a year.

The additional weight of undigested materials does not matter much to aquatic predators, such as fish and crocodiles, but it can be significant for birds like eagles, hawks and owls, which carry prey or catch it in flight. Within hours (minutes in small species) of feeding, owls regurgitate pellets consisting of bones, feathers, hair and other indigestible parts, thus removing unnecessary weight before hunting again.

Almost all animal phyla include some carnivorous species and the range of predation strategies is enormous. Snakes and spiders are among the most specialized and successful predators and have evolved sophisticated ways of subduing prey that are large relative to themselves and of combining predation with digestion.

### 2.2.1 SNAKES

All snakes are strictly carnivorous, feeding mostly on other vertebrates, though many species eat snails, large insects and other invertebrates when immature. Although they have similar diets, and both are reptiles, crocodiles and snakes have very different styles of predation.

What are the major contrasts in the general structure of snake skulls (Figure 2.5) and a crocodile skull (Figure 2.4)?

- The snake skulls appears fragile, obviously unsuited to grasping and holding active prey, or for tearing food. The snake's teeth point backwards and are long and slender. Crocodile teeth are stout and upright.

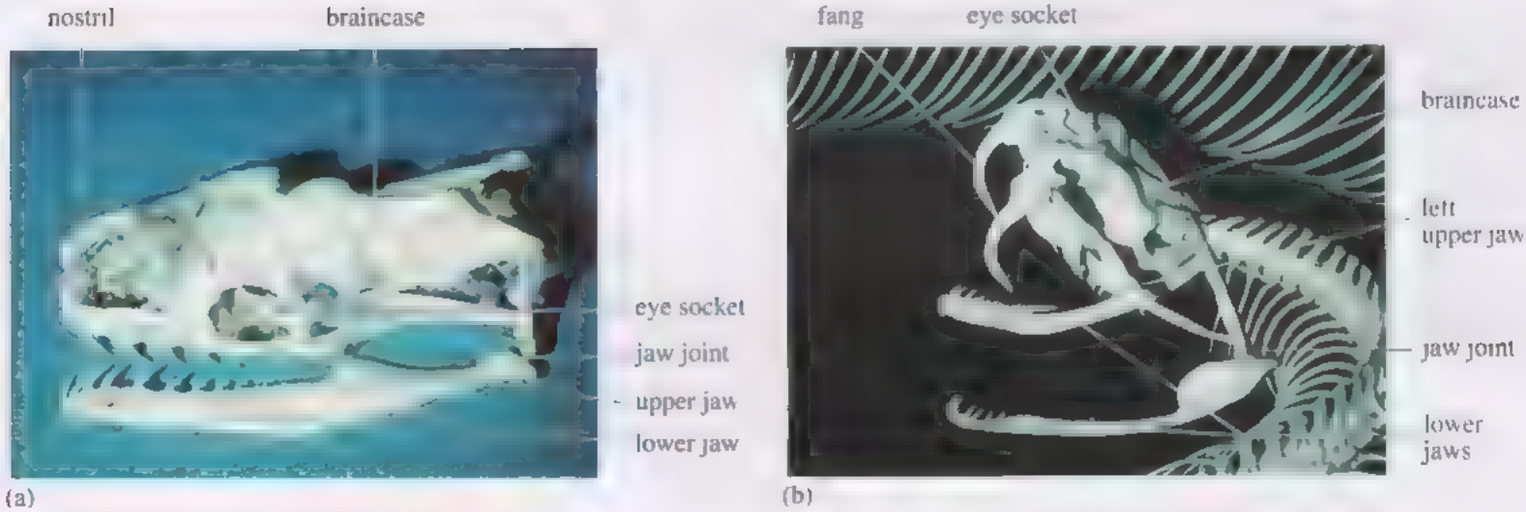



**Figure 2.5** The skull of (a) a python, a large constricting snake and (b) a Gaboon viper (*Bitis gabonica*), a venomous snake. Notice that the jaws are only loosely attached to the skull and can separate from each other, and the teeth point backwards. The fangs are modified teeth on the front part of the upper jaw.

In snakes, the upper and lower, left and right jaws can move relative to each other and the rest of the skull, while in crocodiles, the upper jaws are firmly attached to the nose and braincase, and the lower jaw forms a single rigid structure. Non-venomous boas, pythons and anacondas kill by constriction using the body muscles. The teeth (Figure 2.5a) serve only to grasp the prey and swallow it after it is dead. Venomous snakes such as cobras and vipers have fangs, which are enlarged teeth (Figure 2.5b) attached to bones of the jaw, and can be swung forward to stab prey. They enclose a duct formed from a modified salivary gland which synthesizes and secretes venom, some ingredients of which are listed in Table 2.1.

**Table 2.1** Some major ingredients of snake venoms. The proportions differ greatly between species, depending upon type of prey and predation strategy.

| Toxins | Mode of action | Effect on tissues |
|---|---|---|
| phospholipases | break ester bonds in phospholipids | destroy cell membranes, so cell contents spill out |
| proteases | break up proteins at links between amino acids | destroy enzymes and blood haemoglobin |
| phosphatases | fragment ATP | disable cellular energy supply |
| hyaluronidases | break up extracellular materials | destroy mechanical links between cells in tissues, enabling poison to spread |
| neurotoxic peptides* | bind strongly to molecules that receive chemical messages from neurons | block the transmission of signals from nerves to muscles, causing paralysis |

* Certain families only, including sea snakes.

Venomous snakes wait near nests or feeding areas of small mammals or birds and strike when the prey comes within range, making a small but sometimes quite deep wound into which **venom** from the gland is squirted. The strike may be complete in as little as a few milliseconds, and the snake retreats at once, thereby minimizing scope for the prey to retaliate and injure its predator, and follows the poisoned animal at a discreet distance until it collapses. Field experiments show that snakes can distinguish a stricken individual from others of the same species, probably by smell.

As soon as the prey is no longer able to struggle, the snake swallows it whole. This process often takes many minutes, even hours, as the snake disconnects each half of its flimsy jaws from the rest of the skull, slowly engulfing an object much wider than its own head and forcing it down the throat with its backward-pointing teeth (Figure 2.5). When combined with a sit-and-wait strategy, injecting venom greatly reduces the muscle mass and power required to subdue prey. Many highly poisonous snakes such as mambas and sea snakes are slender and positively delicate compared to the muscular, non-venomous pythons and anacondas.

The neurotoxins (Table 2.1) prevent the prey from struggling, thus making it easier to swallow. Most sea snakes, including the widespread and, in tropical seas, abundant genus *Pelamis*, are slender with small, narrow heads, but their venom is one of the fastest-acting and most lethal of all known natural toxins, paralysing vertebrate muscles in seconds. Sea snakes do not let go of their prey, or do so only very briefly. If they did, the dying prey might sink beyond their reach. But a struggling fish could seriously injure a snake's mechanically weak jaws.

If the dose of neurotoxins was large enough to paralyse their respiratory muscles, terrestrial mammals and birds, and fast-swimming fish, which need lots of oxygen and actively ventilate their lungs or gills, would quickly die from suffocation. But amphibians, which take up oxygen through the skin, and reptiles, especially semi-aquatic species, which breathe infrequently, would take much longer to die from the enzymatic breakdown of tissues. The immediate cause is usually the destruction of blood cells and the lining of blood vessels. Many invertebrates are even less disabled by snake venom.

› Would snakes' style of feeding be effective in avoiding infection by parasites in its prey?

● No. Very little venom reaches most internal parasites and absence of the capacity for mechanical breakdown of the prey means that they enter the snake's gut unharmed. The very slow digestion gives the parasites plenty of time to establish themselves in the snake's body.

Snakes harbour a wide variety of animal parasites, almost all derived from their prey. Compared to other extant groups of reptiles (lizards, the tortoises, turtles and terrapins and the crocodiles and alligators), snakes have quite short lives, rarely more than 30 years. Some connections between longevity and susceptibility to disease are discussed in Chapter 6.

### 2.2.2 SPIDERS

Spiders are among the most ancient and successful terrestrial predators. All of the many thousands of species of spiders are predators, as are many other groups of the class Arachnida (e.g. scorpions, opiliones ('harvestmen'), some mites and ticks). Spiders' main prey is (and so far as we know always has been) other arthropods, especially flying and walking insects, although a few famous spider species are large enough to handle small vertebrates such as birds, hatchling lizards or small fish.

Spiders owe their success in subduing prey that are often much larger and stronger than themselves to the capacity to make **silk**, an unusual protein that forms traps, webs and draglines, and to various poisons which immobilize (but may not always kill) their victims. Both these aids to predation are based on some remarkable protein chemistry and, since the materials are secreted out of the spider's body, they are readily accessible for detailed study. Understanding how they work, and how they can be adapted to thousands of different styles of capturing prey, has involved engineers and crystallographers as well as biochemists and animal behaviourists.

Many insects also produce silk but usually do so only at certain stages of the life cycle, such as to make a cocoon for the pupa, or to attach egg cases to fixed objects. Spiders can produce silk throughout their lives, though its functions may change as the animal matures. Arachnid and insect silks are basically similar, although that of the silk-worm has been much more intensively studied because of its industrial importance.

Silk consists mainly of the small amino acids glycine and alanine, and forms a tightly packed protein consisting of short lengths of stiff 'crystallites', which make the material strong, embedded in a more disordered rubber-like matrix, which confers flexibility, elasticity and resistance to tearing. Stiffer, harder silks contain more, larger crystallites while the matrix predominates in tough, flexible or very stretchy materials, which can be stretched to form continuous very fine strands. Weight for weight, silk is stronger than cellulose or bone, and can be as strong as nylon.

In its chemical structure and mechanical properties, and in the fact that its uses are always external, silk resembles keratin, the protein that forms hair, feathers, horns and claws in terrestrial vertebrates, but differs from it in the way it is produced. Keratin accumulates in the cells that synthesize it until they die, stuck together to form a strand or sheet. Silk is a secretion that, in various different kinds of arthropods, may be produced from modified salivary glands or other parts of the gut, or from reproductive glands. Like all secretions, silk is initially in solution, but the final region of the special ducts through which it is extruded secretes a strong acid, which polymerizes the protein to form an insoluble, biochemically inert material. Web-building spiders build by running round the web as shown in Figure 2.6a, at a pace that matches the rate of extrusion of the silk, thus stretching it to form a very thin thread as it toughens.*

The first web-building spiders appeared in the early Devonian 400 Ma ago, some probably caught walking prey by means of trip-lines set from small holes in the ground, while others spun three-dimensional cotton-wool-like masses which entrapped animals that fell onto them. Both these methods of catching prey are still found among living species, but the most sophisticated, and most recently evolved, kind of web is the flat, two-dimensional 'orb web' which catches flying insects. One of the most familiar orb weavers in Britain is the garden cross spider (*Araneus diadematus*), which is common in woodland, pasture and gardens, where its varied diet includes flies, bees, wasps, aphids and beetles, depending on location and season. In captivity, it thrives on meals of *Drosophila* and young locusts or crickets, and has thus been the subject of detailed laboratory investigation.

Catching a fast-flying insect on a flat web is comparable with trying to stop, and hold, a ballistic missile with a wire fence, and requires some subtle engineering. *Araneus* can produce at least seven kinds of silk, each of slightly different chemical composition, which are synthesized and secreted from structurally distinct glands, together occupying a large fraction of the bulbous abdomen. The spider uses different silks for the radial and spiral threads of the web (Figure 2.6a), a third kind for binding these two strands together and for the attachments to twigs or other suitable moorings, and yet more kinds for roles such as wrapping prey and as climbing lines.

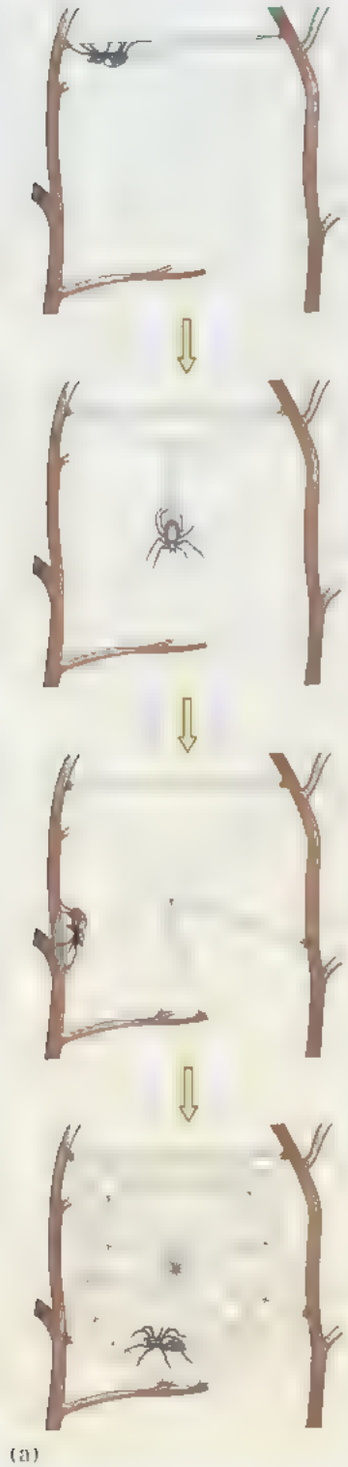
As it is extruded, the silk used to form the spiral strands is coated with a viscous mixture of glycoproteins (proteins with sugar side-chains) and certain free amino acids secreted from yet another gland. The glycoproteins readily form weak chemical bonds with other substances, including water, so the mixture is hygroscopic (water attracting) and sticky. It takes up water vapour from the air, forming sticky globules that are large enough to be visible in bright sunlight as tiny 'beads'. Like many fibres (e.g. wool), this type of silk is much more elastic when damp, enabling it to absorb the energy of an insect hurtling into the web.

* Spiders are often said to 'spin' silk thread. This term is avoided here because the mechanics of silk extrusion have little in common with 'spinning' yarn from wool or cotton and more closely resemble making wires or cables.

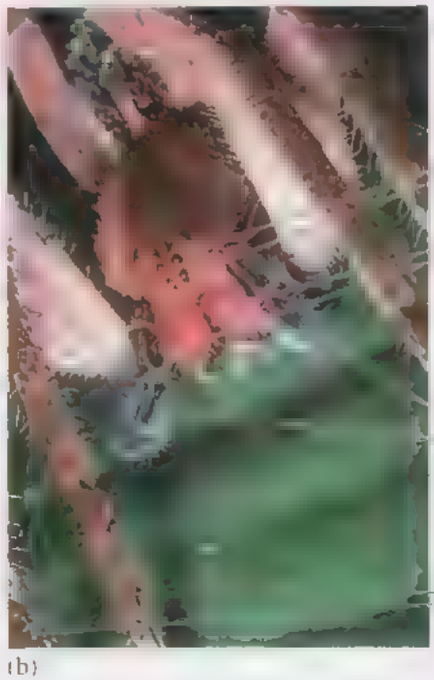
The impact stretches the threads to up to three times their normal length, but the energy is stored elastically, so the web recoils quickly to its original shape. The capacity to absorb impact in this way makes the web durable when built in places where it is most likely to intercept plenty of prey. Fast-flying insects such as bees, and wind-borne debris, may punch holes in the web or rip it from its moorings, thereby disabling it and dispersing its materials.

The stickiness conferred by the glycoproteins helps to hold the prey to the web, otherwise the insect would bounce off, like a person landing on a trampoline, and be lost to the spider (sometimes prey do escape in this way). Although densely woven, sticky webs catch prey more efficiently, they are also more resistant to air currents so are more likely to be torn by strong winds. Many spiders that thrive in a variety of habitats adjust their style of web-building to the weather. Experiments on captive *Araneus* show that under moderately windy conditions, spiders build smaller, rounder webs than they do in still air, and they preferentially place them parallel to, rather than perpendicular to, the direction of the prevailing wind. Orb webs cannot be used at all in exposed positions in very windy climates. *Araneus* webs are largest and most abundant, and hence most easily noticed, in sheltered woodland and gardens, and inside buildings.

A substantial amount of the spiders' body mass is cuticle and other non-reclaimable tissues, so making a web represents a significant outlay of protein. *Araneus* replaces its web about once a day during the summer hunting season, but other species build massive, elaborate webs that last for weeks. If spiders are induced to build new webs more frequently than normal, they use less and less silk each time, suggesting that the rate of silk synthesis limits web-building capacity.



(a)




(b)

**Figure 2.6** (a) Stages in the construction of a typical orb web. The spider uses different silk for anchoring the web (top), the radial threads (middle) and the spiral threads (bottom). (b) Photograph of a large tropical spider showing the pedipalps and chelicerae, which are paired feeding appendages in front of the very small ventral mouth (not visible). See text for details.

1 From the information in Table 2.2, how long does it take to make a new web?

● The total length of the thread is 14 + 4 = 18 m, which is produced at an average rate of 1 cm $s^{-1}$, or (60 × 60)/100 m $h^{-1}$ = 36 m $h^{-1}$. To produce 18 m at 36 m $h^{-1}$ takes 0.5 h, so a new web is made in about 30 min.

**Table 2.2** Some average properties of orb-weaving spiders and their webs. Data from Vollrath *et al.* (1997).

| | |
|---|---|
| Body mass | up to 50 mg |
| Number of radial threads | 33 |
| Diameter of silk thread | 1–2 μm |
| Total length of radial threads | 4 m |
| Total length of spiral strands | 14 m |
| Web area | 400 $cm^2$ |
| Mass of web | 0.1–0.5 mg |
| Rate of extrusion of silk | 1 cm $s^{-1}$ |

Web building is quite strenuous, because the spider runs around the web to lay its thread. Only a small fraction of the day is spent in this activity. For the rest of the time, spiders maintain mechanical contact with their web, usually waiting near the centre or at an attachment point. Their legs (see Figure 2.6b), like those of almost all arthropods, are very sensitive to vibration. As soon as the spider feels the impact of an insect on its web, it scrambles towards the disturbance and immobilizes the prey before its struggles do too much damage to the delicate fabric.

? From the information in Table 2.2, what percentage of the total body mass is used to make a new web?

● Spiders weighing 50 mg secrete up to (0.5/50) × 100 = 1% of the total body mass as silk to make each web.

Under ideal conditions, spiders catch prey amounting to about 3–12% of their body mass each day. Many spiders, including *Araneus*, eat their own web when it becomes too tattered to be effective, thereby reclaiming much of the protein they put into it. The water absorbed from the air by the glycoprotein glue, or collected as dewdrops, is a bonus, as spiders do not normally drink. The soft tissues of their prey are the only other source of water, and their respiratory and excretory mechanisms are well adapted to conserving water.

? Why can such recycling of the silk not sustain web-building indefinitely?

● Metabolic energy is used in producing enzymes that digest the silk to its component amino acids in the gut, in transporting them from the gut into the bloodstream, and in re-using them for synthesizing more silk, or another kind of protein.

If their webs fail to net any edible prey, the spiders' body reserves are gradually depleted until they are unable to support further web building. Without its means of capturing prey, a webless spider starves to death, often leaving its final web unreclaimed. Like other opportunist predators, most spiders can fast for a remarkably long time, over 200 days in the case of adult black widow spiders. Their rate of using energy falls by as much as 80% and the spiders' tissues are slowly depleted of glycogen, fats and finally proteins.

Larger, more conspicuously placed webs have a better chance of catching prey, but they also use more silk, and are more susceptible to damage. Establishing a satisfactory ratio of the energy outlayed to collect food and the nourishment so obtained is called **optimal foraging**. In the case of orb weavers, predation is most efficient when the silk remains strong when stretched into a fine thread, thus minimizing the quantity of material used, and there is a good match between the chemical and mechanical properties of the web and how and where the spider deploys it.

As you would expect, several lineages of insects have evolved countermeasures that help them avoid entrapment in spiders' webs. Among the most familiar are the fine, loose, often highly coloured scales on butterflies' wings. If a wing touches a newly formed web, the butterfly is often only briefly disorientated, because its scales stick to, and thereby inactivate, the web, enabling the insect to escape with only minor damage. Sticky webs also accumulate fine particles of dust. Except when coated with dew on a sunny morning, we do not usually notice orb webs until they become so covered in dirt as to be useless to the spider. Such tough, dry protein is only slowly degraded by bacteria and fungi, so unless fragmented by wind or rain, old spiders' webs accumulate, especially in sheltered corners of unused buildings.

The use of webs has made spiders by far the most important predators of insects on the wing, competing only with a few species of birds, such as swifts, swallows and flycatchers, insects such as dragonflies, damselflies and certain wasps, and insectivorous bats. With such subtle relationships between web structure, durability and function, it is not difficult to see how spiders have diversified in structure and habits, each species producing webs and draglines suited to different conditions of weather and terrain, and catching different kinds of prey.

Most spiders wrap captured prey in silk, then inject venom through the powerful, stabbing appendages called chelicerae (see Figure 2.6b). The poison glands are squeezed by muscles, enabling the spider to control how much is administered. 13 ng (1 ng = $10^{-9}$ g) of the poison of black widow spiders is sufficient to kill a housefly, and about 15 µg (1 µg = $10^{-6}$ g) disables a large cockroach within 10 minutes. Most spider venoms contain toxins that interfere with aspects of nerve transmission that are specific to arthropods and do not affect vertebrate neurons, so only a few species are dangerous to humans. The spider may eat its prey at once, or enfold it in a mesh of silk and store it alive but paralysed on or near its web for later consumption.

) What is the advantage to the predator of keeping the prey alive?

- Although it cannot move, the prey's cellular mechanisms that protect its body from invasion by bacteria and fungi continue to operate, thereby preventing the normal processes of decay.

The capacity to store prey in this way enables spiders to take maximum advantage of transient gluts of food without greatly increasing their rate of feeding or size of internal storage tissues.

Some predatory insects, notably digger wasps, also use venoms that paralyse but do not kill their prey. This strategy works much better with terrestrial arthropods whose respiratory organs can operate without muscular activity, than with vertebrates whose lungs or gills must be actively ventilated and the blood pumped around the body. Enough oxygen diffuses through arthropods' respiratory surfaces to vital tissues to keep them alive as long as they remain sedentary, but when a vertebrate's muscles are paralysed, it dies from suffocation, usually within a few minutes.

Spiders feed by squirting digestive enzymes through the mouth onto the immobilized prey, then using the muscular gut to suck up the liquidized tissues through fine filters around the mouth and in the anterior part of the gut. All particles larger than about 1 μm are excluded, and the filter is cleaned from time to time by spitting out digestive juices. Tough cuticle and other parts of the prey that cannot be digested often remain on or near webs and may be sufficiently intact to be recognizable.

Because so much digestion occurs outside the body, the spider gut is quite short, though the glands that produce the digestive enzymes can be extensive. These enzymes accumulate in granules inside secretory cells in the gut lining, and can be expelled into the gut within a few minutes when required. Other absorptive cells take up the digested material into vacuoles.

❍ Would spiders produce much faeces?

● No. Sucking in partially digested prey avoids taking in indigestible material such as cuticle.

The small quantity of faeces is mainly material from the gut itself.

❍ Compared to vertebrates, how would this style of feeding affect susceptibility to pathogens in the prey?

● Sucking in the liquefied remains of externally digested prey avoids contact between the gut lining and many pathogens, reducing the risk of infection compared to vertebrate predators that ingest the prey whole or in large chunks.

External digestion of prey similar to that of spiders is common among carnivorous insects and other invertebrates.

## 2.3 HERBIVORY

Consumers of terrestrial plants face very different problems from those of predators. Plants are often widespread and abundant but much of the tissue is indigestible, each cell being surrounded by a tough cell wall. The nutritious components can be reached only by piercing through the cell walls or by mechanically crushing the cells to release their contents, strategies that require hard tissues such as toughened cuticle or teeth, operated by powerful, well controlled muscles, or by chemically digesting away the tough protective materials. As well as this mechanical protection, the leaves and roots of many plants, especially large, long-lived species, also synthesize various toxins that deter herbivory by destroying or impeding animals' secretions or intracellular processes.

) How could herbivorous animals deal with toxins distributed throughout the plant?

● They would be unable to avoid ingesting toxins along with the nutritious tissues, so their digestive system has to prevent them being absorbed, or detoxify them and excrete the waste.

In view of these fundamental problems of eating living plants, it is not surprising that, although almost all phyla and major classes of animals include predatory species, only a few lineages have become as diverse and widespread as terrestrial herbivores. Most terrestrial slugs and snails (phylum Mollusca, class Gastropoda, order Pulmonata) rasp leaves and stems, though many species are limited to foraging in damp habitats or at night. Many roundworms (phylum Nematoda) burrow into plant roots and other underground structures such as bulbs and tubers, sucking out their nutrients.

By far the most ancient and numerous consumers of living plants are the arthropods, including millipedes and above all insects, which comprise the great majority of terrestrial herbivores. Major orders that consist almost entirely of herbivores include Isoptera (termites), Orthoptera (locusts, grasshoppers and crickets) which bite and chew leaves and stems, Hemiptera (cicadas, shield bugs, leaf hoppers, spittlebugs, aphids, whiteflies) which feed by piercing and sucking plant sap, and Lepidoptera (butterflies and moths) of which most larvae (caterpillars) nibble leaves, and many adults take nectar. Most earwigs (Dermaptera) and many families of Coleoptera (beetles), including chafers, weevils, leaf beetles and woodworm, are also obligate herbivores, and many Diptera (flies) and Hymenoptera (ants, bees and wasps) depend on plants during part of their life cycle.

Jawed vertebrates are now, and have been throughout their evolutionary history, primarily predators, though a few species of most classes have become herbivores. Among living vertebrates, a few fish (including goldfish) eat free-floating or encrusting algae and/or aquatic plants such as pondweed. Most tortoises and a very few lizards (iguanas and their relations) eat leaves and flowers, at least when fully grown, though almost all species supplement their diet with snails, worms, insect larvae or carrion if they find them. Many kinds of birds (e.g. parrots, pigeons, finches, grouse, pheasants) feed on concentrated plant tissues such as seeds, buds or fruits, but only a few, notably ostriches and their relatives, and geese, plus some ducks and moorhens, eat mature leaves in large quantities. Throughout the Tertiary, the most important vertebrate consumers of foliage have been the mammals, mainly because they can chew very efficiently.

) From your own experience, what features of the mouth and teeth are necessary for chewing?

● Chewing requires broad, fairly flat teeth mounted on exactly opposable jaws so the teeth 'fit' together to form a grinding surface. The process also entails being able to breathe while the mouth is full, and manipulating food around the mouth with the tongue and cheeks.

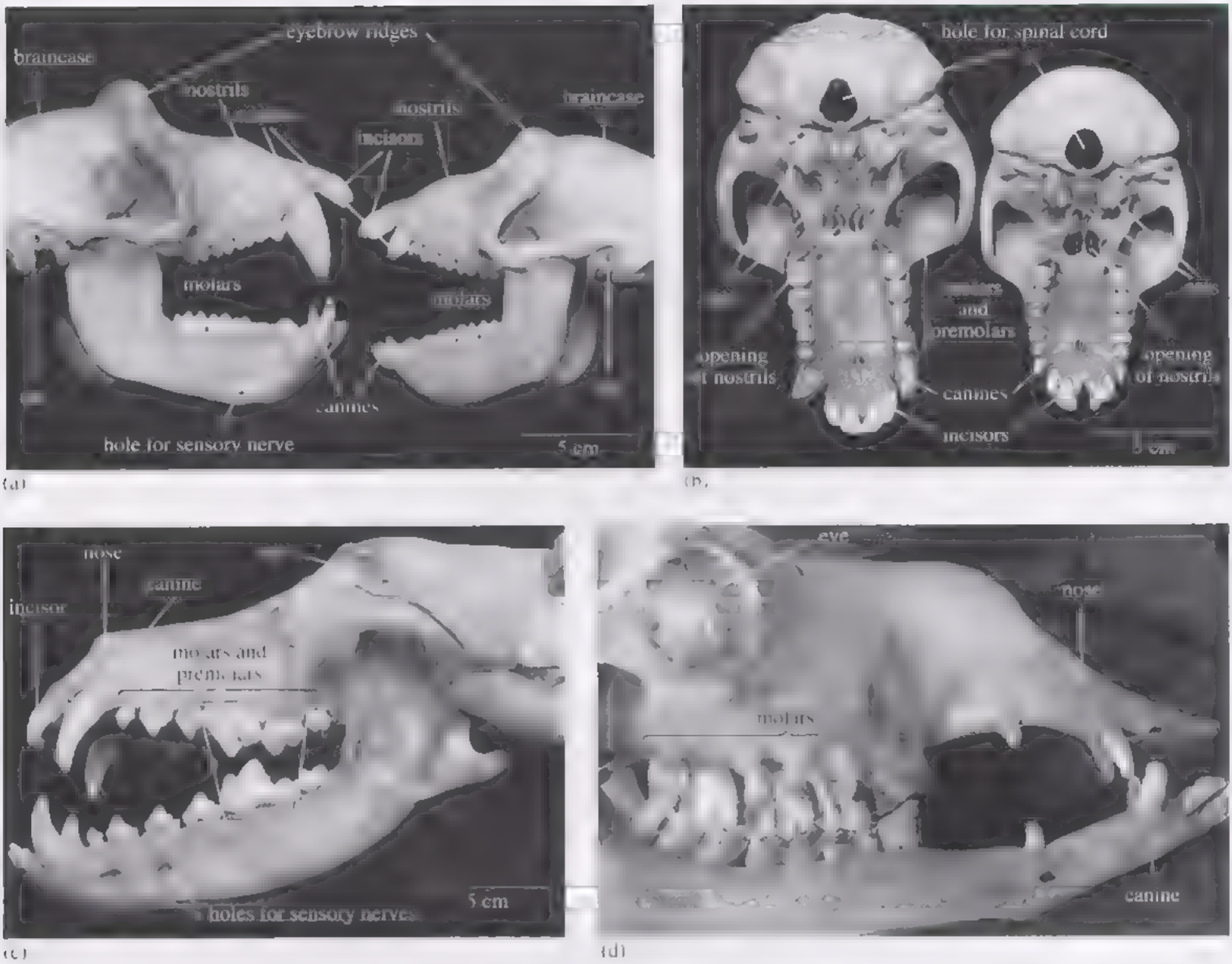



**Figure 2.7** Some mammalian skulls. (a) and (b) The skulls of an adult male (left) and female (right) macaque monkey, showing the main features of mammalian dentition. The opening of the nostrils into the throat at back of the mouth can be seen in (b). (c) The skull of a timber wolf, a specialized carnivore, with cutting molars outlined. (d) The teeth and jaws of a camel, showing the grinding molars and the exact fit between the upper and lower teeth

) Comparing Figure 2.7 with Figures 2.4 and 2.5, how do mammalian teeth differ from those of reptiles?

- The mammals have fewer teeth than reptiles, but they are complex in structure, highly differentiated and those of the upper and lower jaws are exactly opposed.

With a few exceptions (notably the marine mammals), the dentition consists of front, gripping incisors, piercing canines and grinding molars and premolars, as shown in Figures 2.7a and b. In carnivores, the dagger-like canines are massive and the stout, pointed molars are adapted for shearing (Figure 2.7c). The canines

are reduced in most herbivores,* which nibble, tear or gnaw plants between their incisor teeth and/or lips, then grind each mouthful between their broad, hard, elaborately sculptured molar teeth (Figure 2.7d). Chewing breaks insoluble plant tissues into tiny fragments which form a suspension in the copious saliva, maximizing the surface area on which enzymes can work.

Digestion of plants is slower and usually less efficient than that of animal tissues, so nearly all herbivorous animals have more extensive guts than their carnivorous relatives. Most herbivorous vertebrates rely upon some form of microbial-assisted digestion. The guts of foliage-eating mammals (and those of a few herbivorous birds) have special anatomical structures that enable them to harbour huge numbers of symbiotic microbes. In 'hindgut fermenters', the colon, and/or the **caecum**, a blind-ended sac extending from the junction of the small and large intestine, are greatly enlarged and harbour symbiotic microbes. The gut contents thus reach the microbes after the animals' own enzymes have digested them as much as they can. In other groups of herbivores, the microbes are concentrated in forestomachs, which form as a pouch in the oesophagus, anterior to the true stomach. The lining of forestomachs does not produce secretions and does not actively take up nutrients. The microbes thus have access to the food before the animals' own enzymes reach it, which has profound implications for the mammals' diet, nutritional requirements and metabolism.

## SUMMARY OF SECTIONS 2.2 AND 2.3

1 Animals can prey upon each other in a huge range of ways, and predators have a variety of habits and body form.

2 Primitive vertebrates, such as crocodiles, stalk prey and then seize it with stout teeth, tearing and shaking it with their powerful jaws and body. Large chunks are swallowed whole.

3 Poisonous snakes inject venom into prey. It contains enzymes that break down tissues and red blood cells, and, in more advanced families, neurotoxins. The fangs are modified teeth. The jaws are mechanically weak but can be disarticulated to swallow immobilized prey whole. Digestion of large prey may take weeks.

4 Spiders build webs which trap flying or walking arthropods by secreting several different kinds of silk and associated materials. Web building requires a significant outlay of body reserves, which cannot be sustained unless enough prey are caught and eaten.

5 The main consumers of terrestrial plants are insects, gastropod molluscs, nematodes, a few birds and many kinds of mammals, whose guts are modified to harbour microbial symbionts.

* Canine teeth are enlarged in adult males of certain primitive deer that lack antlers, including the Chinese water deer (*Hydropotes inermis*) now living wild in parts of southern Britain. They function in social and sexual behaviour.

## 2.4 DIGESTION

Digestion is the chemical breakdown of ingested food. Organic material is attacked by **digestive enzymes**, while inorganic matter (the mineral components of shells, bone, etc.) is dissolved by acids. The main sites of digestion are the stomach and anterior part of the intestine, though in many animals, the saliva also contains some enzymes, enabling digestion to begin in the mouth. The main classes of digestive enzymes are **proteases** * which fragment proteins by hydrolysing peptide bonds between amino acids, **lipases** which hydrolyse the ester bonds that link fatty acids to glycerol in membrane phospholipids and storage triacylglycerols, and various enzymes that break down complex carbohydrates into monosaccharides.

Digestive enzymes are basically similar in all heterotrophic organisms, bacteria and fungi as well as the huge range of animals, though species differ in the amounts of the different kinds of enzymes they produce, and in the chemical conditions (e.g. pH, temperature) under which they operate, depending upon their diet and habitat. Those produced by carnivorous plants such as sundews (*Drosera*), which digest insect prey, are remarkably similar in molecular structure and mode of action to the corresponding secretions found in animals, even though other green plants do not break down complex organic molecules to obtain nutrients.

Digestive enzymes are usually less specific than intracellular enzymes, acting, with various degrees of efficiency, on a wide range of similar molecules, especially in omnivorous organisms which eat many different foods. Those of vertebrates tend to be more specific than those of fungi and invertebrates, both for the kinds of bonds they attack, and the chemical conditions under which they work most efficiently, and so they often occur in greater variety. For example, of the proteases, **pepsin** is secreted by the stomach epithelium; it works best in strongly acidic solution and cleaves peptide bonds formed by amino acids in which the R group is negatively charged, i.e. acidic (e.g. aspartate), or has a ring structure (e.g. phenylalanine). **Trypsin** is secreted into the small intestine by the **pancreas** gland. This enzyme functions at or near neutral pH and preferentially attacks bonds involving the amino acids arginine or lysine, both of which have positively charged (basic) R groups. The variety is further increased by the fact that some proteases, including pepsin and trypsin, can cleave peptide bonds in the middle of a polypeptide chain, but others work preferentially on the N- or C-terminal amino acids. The combined action of these enzymes can reduce a wide variety of proteins to single amino acids or small peptides, molecules that are small enough to be conveyed through the absorptive epithelium of the intestine.

- How could bonds that are usually broken by digestive enzymes escape hydrolysis?

- Breakdown could be very slow if the higher order structure (i.e. folding) of the molecule were such that the enzymes could not get close enough to the relevant chemical bonds.

* Some, but not, all enzymes are named after their substrate, e.g. protease (proteins), lipase (lipids), cellulase (cellulose), lactase (lactose).

The acidity of the stomach helps to open up tightly folded proteins, thus facilitating their digestion, but most vertebrates cannot break down tough materials such as keratin (see Section 2.2.2), silk or tooth enamel fast enough to yield useful amounts of amino acids. These molecules may pass right through the gut with little or no digestion.

As well as immobilizing the prey, snake venom also contributes to digestion.

› Which components of venom listed in Table 2.1 aid the digestion of the prey?

● The phospholipase, protease and hyaluronidase enzymes.

The combined action of these injected enzymes which break down the tissue from the inside and secreted enzymes attacking it from the outside speeds up the snake's digestion. But the process is still slow compared to that of animals that shred or chew their food before swallowing it. Even under ideal, warm conditions, snakes can take weeks to complete the digestion of a large meal

› Would the neurotoxin component of the venom help to digest the prey?

● No. Neurotoxins bind to proteins, rather than enzymatically break them down, so they contribute nothing to digestion.

Detailed studies of the amino acid sequences of enzymes reveal that many of those secreted for digestion are similar to the enzymes that perform analogous transformations in other tissues. For example, pancreatic lipase which, in the vertebrate intestine, breaks down triacylglycerols and phospholipids in seeds and in the adipose tissue of prey (see Section 2.4.2), proves to be only slightly different from the lipoprotein lipase found in muscle, adipose tissue and many other tissues which hydrolyses blood-borne triacylglycerols, releasing fatty acids that these tissues take up and use themselves (see Section 2.6.2).

› What changes in gene activation could have led to the evolution of venom production (Section 2.2.1)?

● As shown in Table 2.1, snake venom consists mainly of enzymes that in other vertebrates are involved only in digestion. The genes for these proteins are present in all cells, so the major evolutionary change was their activation in the salivary glands as well as in the lining of the gut and/or in the pancreas.

The evolution of venom production is thus probably quite simple, and indeed comparative studies suggest that it has happened many times in different lineages of animals. The adaptations of the teeth and jaws (Figure 2.5b), not to mention those for detecting and striking at prey, are more complicated, and are unique to snakes. The salivary glands of the Gila monster (a predatory lizard found in the deserts of southwest USA) also produce venom, but poisonous secretions are rare among birds and mammals, being known from only a few species, notably shrews. Other large lizards, the Varanidae, harbour many potentially dangerous bacteria in the mouth, so a shallow bite on a limb is very likely to turn septic. The swelling and delayed healing lames the potential prey, which is much easier to catch while it is partially disabled.

### 2.4.1 ADJUSTING THE ENZYMES' WORKING ENVIRONMENTS

All enzymes have a narrow range of concentrations at which they work most efficiently. To maximize the speed and thoroughness of digestion, the dilution of the gut contents is adjusted by removing or adding water. Sea-anemones and their relatives *Hydra* have one of the simplest means of doing so: the body retracts after prey is caught, expelling much of the water in the enteron (gut), and remains closed for hours or days, thus trapping the food and digestive enzymes together (Figure 2.8).

Another means of preventing loss and excess dilution of digestive enzymes is phagocytosis, which is followed by intracellular digestion. In heterotrophic protoctists, including the unicellular *Amoeba*, the products of digestion diffuse into the rest of the cell, and any insoluble waste is expelled as the vacuole reaches the surface. Similar processes occur in the cells lining the guts of many kinds of invertebrate animal, especially particle-feeders.

Terrestrial animals usually need to add water to the food before digestion begins, turning the mixture into a slurry. The gut epithelium or associated glands (e.g. the salivary glands of insects and vertebrates) may secrete large quantities of water, which persists in the stomach and intestines and is reabsorbed, together with salts and other small molecules, in the hindgut after digestion and absorption are complete. The faeces of terrestrial reptiles, birds and mammals are thus comparatively dry. Terrestrial insects also reabsorb most of the water in the gut contents.

The mechanisms and cellular conditions by which water passes into and out of the gut are explained later. The important point is that they require energy and may entail excessive loss of ions in body fluids. Extracting water from the faeces contributes substantially to the metabolic cost of digestion, so the process is less thorough (i.e. the faeces remain wetter) in animals with easy access to drinking water than in desert-adapted animals.

› Would aquatic animals have a well-developed hindgut?

● No. There is little active reabsorption of water where plenty is available to the animal.

Among vertebrates, fish have no distinct hindgut. In terrestrial reptiles, birds and mammals, the water-absorbing colon is clearly differentiated from the nutrient-absorbing small intestine.

Some animals, including a few land-dwelling forms, have the opposite problem: to remove excess water from dilute foods. As well as diluting enzymes, water also adds to the body mass, which does not matter for aquatic (i.e. neutrally buoyant) or sessile species, but can be a significant problem for flying animals, such as insects. Special physiological adaptations are needed to enable bees to utilize nectar, which is nutritious and contains no toxins, but is too dilute to be digested efficiently.

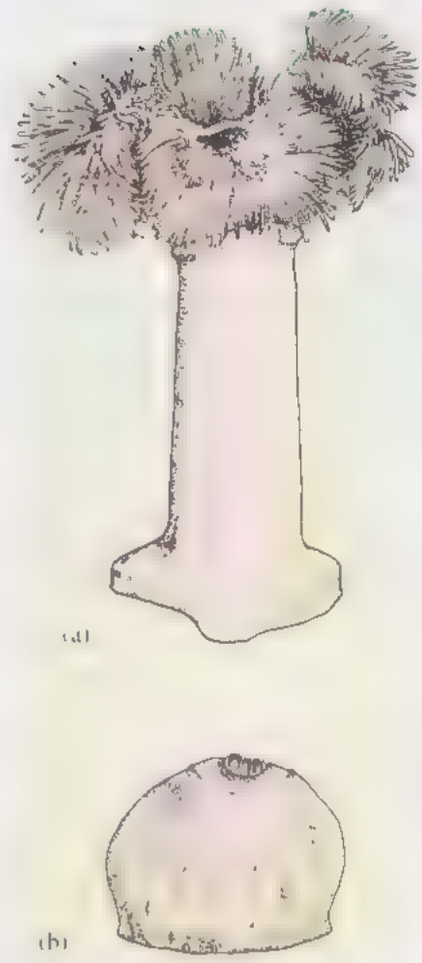



**Figure 2.8** A sea-anemone (*Metridium senile*, phylum Cnidaria) (a) expanded with tentacles protruded to collect food and (b) closed up while digesting food.

### BUMBLEBEES

Although sedentary as larvae, adult bumblebees (*Bombus lucorum*) fly to find mates and suitable places in which to lay eggs. Flight requires large quantities of metabolic fuel which, in the case of bumblebees, comes almost entirely from nectar, a concentrated solution of fructose, sucrose and other simple sugars in water. *Bombus* flies from flower to flower, sucking in 0.01–0.1 µl of nectar at each visit (depending on the species).

In a detailed study of captive bumblebees fed from artificial 'flowers' that provided about 1 µl of 'nectar', males that weighed about 220 mg foraged for about 4 h a day, flying an average of 17 km to collect about 220 mg of 50% sugar solution (i.e. 110 mg of sugar in 110 mg of water). Complete oxidation of 110 mg glucose yields about 66 mg of water, which together with the 110 mg of water in the nectar, makes a total of 176 mg. Only 40 mg of water is lost by evaporation, mostly through the respiratory organs (spiracles). So to maintain constant body mass, about 136 mg of water must be expelled as urine each day. That might not sound a lot, but it represents 62% of a bumblebee's total body mass, and so adds significantly to the energy needed to power flight. The bees cannot avoid taking in this water, but some physiological and behavioural modifications, summarized in Figure 2.9, reduce its impact on their foraging efficiency.

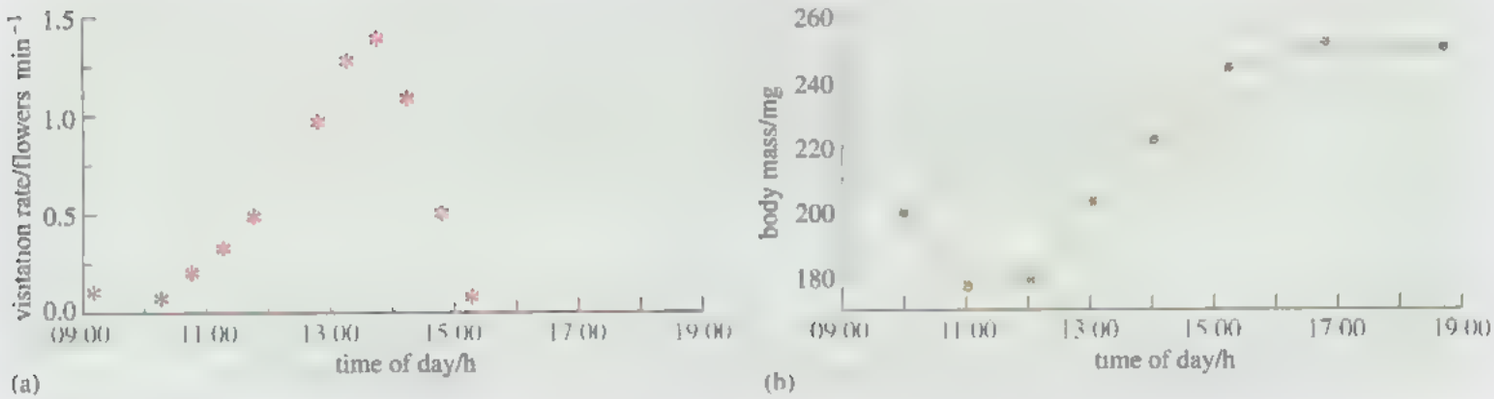



**Figure 2.9** Graphical summary of (a) average frequency of visiting flowers and (b) changes in body mass of male bumblebees between 9 a.m. and 7 p.m. Data from Bertsch (1984)

) What is the relationship between time spent flying and body mass?

● The bees fly most during the time when the body mass is well below maximum.

The body mass falls to a minimum of 28% (= (250 − 180)/250) below its value during the previous night, due to oxidation of sugars stored from the previous day's foraging, and to the excretion of excess water. Nearly all the synthetic 'nectar' was taken in over 3 hours during the warmest part of the day, and when the body mass had increased by about 20% of the minimum value measured at 9 a.m., the bees landed at a suitable roost and rested.

Nectar-feeding bats and humming-birds face similar problems, but the energetic cost of flying is lower, and the rate of water loss through the lungs is higher, for vertebrates than the corresponding values for insects, and their kidneys can excrete excess water more efficiently. So these larger animals can live on nectar that contains as little as 25% sugar, too dilute to serve as a food source for

bumblebees, which are limited to plants that produce nectar with 35–50% sugar. Blood is another dilute solution of useful nutrients; the excretory mechanism of blood-sucking insects such as mosquitoes is so efficient that the excess water is expelled within a few minutes, often while sucking is still in progress, thus allowing for a quick getaway should their host try to swat them.

Whether a food forms an adequate diet for a particular species depends not only upon the nutrients it contains, but also upon how they are presented and the physiological capacities of the eater. This example was chosen because it is particularly clear-cut and easy to quantify, but the principle it demonstrates applies to many other plant–herbivore and predator–prey relationships. Apparently irrelevant abilities, in this case the rate of water loss and flight energetics, determine food choices and diet composition, and hence where and when species can live.

### 2.4.2 Digesting waxes

Digestion of proteins and carbohydrates takes places in aqueous solution but for lipases (lipid digesting enzymes) to work fast and efficiently, lipids must first be freed from the cells that contain them, and emulsified (i.e. mixed with water). In most vertebrates, lipid digestion does not begin until food reaches the small intestine, after the cells have been ruptured by biting or chewing, churning of the stomach and the action of proteases. After they leave the stomach, the gut contents are mixed with **bile**, a secretion of the liver which (at least in most vertebrates) accumulates in a pouch called the gall bladder before being released into the small intestine as required. Bile emulsifies lipids, including triacylglycerols, phospholipids and waxes, thus enabling the lipases secreted by the pancreas to bind them and hydrolyse their ester bonds.

Samples of bile and pancreatic secretions can be easily obtained from freshly killed animals, and their chemical composition and capacity to digest lipids measured *in vitro*. The volume of bile secreted and the concentration of salts in it depend greatly on the diet, being more than 20 times higher in birds such as storm petrels, *Oceanodroma leucorhoa* (Figure 2.10a) and Gentoo penguins, *Pygoscelis papua* (Figure 2.10b) than in chickens or pigeons. These marine scavengers and predators eat krill (large shrimp-like crustaceans) and other planktonic invertebrates in the cold Southern Ocean.

Many surface-living animals contain lipids which make them more buoyant and/or act as an energy reserve, but those that live in cold waters usually have a greater proportion of waxes and a smaller proportion of triacylglycerols, because the latter tend to solidify at low temperatures (like cooking oil kept in the fridge). Waxes mix with water even less readily than triacylglycerols, so they can only be emulsified in more concentrated bile. Conversely, the more dilute bile and pancreatic lipases of seed-eating chickens and pigeons digest triacylglycerols much faster and more efficiently than waxes.

As well as these biochemical studies, the development of new computer-based techniques for scanning *in vivo* has confirmed what has long been suspected: reverse peristalsis, propelling the gut contents back and forth, is frequent in bird intestines, especially those equipped to digest waxes and other difficult foods.

(a)

(b)

**Figure 2.10** (a) Storm petrels cruise just above the waves on long, narrow wings and skim off small floating invertebrates with their long beaks. (b) Gentoo penguins breed near the coasts of islands in the Southern Ocean and Antarctica, where they swim through surface waters catching krill (large swimming crustaceans) and small fish.

The capacity for digesting waxes greatly increases in animals whose dietary habits require it. The genes for the relevant enzymes and modulation of bile secretion mechanisms are probably present in most or all birds, but have become strongly active only in those species whose diet is rich in waxes. Comparative studies of digestion indicate that this principle is generally true: adjustments to digestion parallel changes in diets. Concentrations of secretions and activities of enzymes can differ enormously in animals adapted to different diets.

Among terrestrial organisms, thick masses of wax occur only in certain special structures such as honeycombs, but most green plants and terrestrial arthropods have a thin waxy layer on their outer surfaces which aids waterproofing. Most leaf-chewing insects (e.g. caterpillars) and herbivorous or insectivorous vertebrates can digest small quantities of waxes in this form. This capacity is greatly enhanced in the caterpillar of the wax moth, *Galleria mellonella*, which lives in honeycombs and is a common and sometimes serious pest in commercial honeybee hives. *Galleria* digests waxes and utilizes the resulting fatty acids and alcohols so efficiently that, although almost all other caterpillars obtain most of their energy from the sugars and starches in the plants they eat, this species can be bred successfully on a carbohydrate-free diet.

### 2.4.3 RUMINANTS

Ruminants are the most abundant and diverse mammalian consumers of grasses and leafy shrubs. Their first and largest forestomach (anterior to the true stomach, Section 2.3) is the **rumen**, a huge warm vat of microbes whose activities are greatly facilitated by the habit of **chewing the cud**. The 'true ruminants' are advanced families of artiodactyls, including giraffes, pronghorns, antelopes, gazelles, goats, buffalo, bison, cattle, sheep, and most kinds of deer, which have four stomachs, and camels, llamas and a few primitive deer, which have three.

Ruminants take numerous small bites in quick succession, with only a minimum of grinding between the molar teeth. The coarsely fragmented plant material is combined with large quantities of saliva (several hundred litres per day in the case of cattle) and swallowed into the rumen (Figure 2.11a and b), where powerful muscular churning mixes it with symbiotic bacteria, fungi and protoctists (mostly ciliates) in combinations that depend greatly upon the species of plants the animal habitually eats. Larger particles of plant material float to the top of the rumen and pass into the reticulum (Figure 2.11b), a much smaller chamber whose function is primarily mechanical, forming boluses of cud which are regurgitated for further chewing. At intervals from minutes to hours (usually longer in larger species), ruminants pause to chew the cud, often lying down in groups to do so. Morsels of cud pass back to the mouth where they are ground between the large molar teeth by slow, powerful side to side movements of the jaws. The cud is re-swallowed and, if its particles are fine enough, sinks into the rumen where it is fermented by microbes over a period of from a few hours to several days, depending on the quality of the forage.

Efficient fermentation requires the microbes to adhere well to the plant fragments so that the enzyme concentration at the exposed surface is as high as possible. Adhesion is assisted by thorough chewing, which greatly increases the surface area to which the symbionts can attach, but also depends upon the chemical

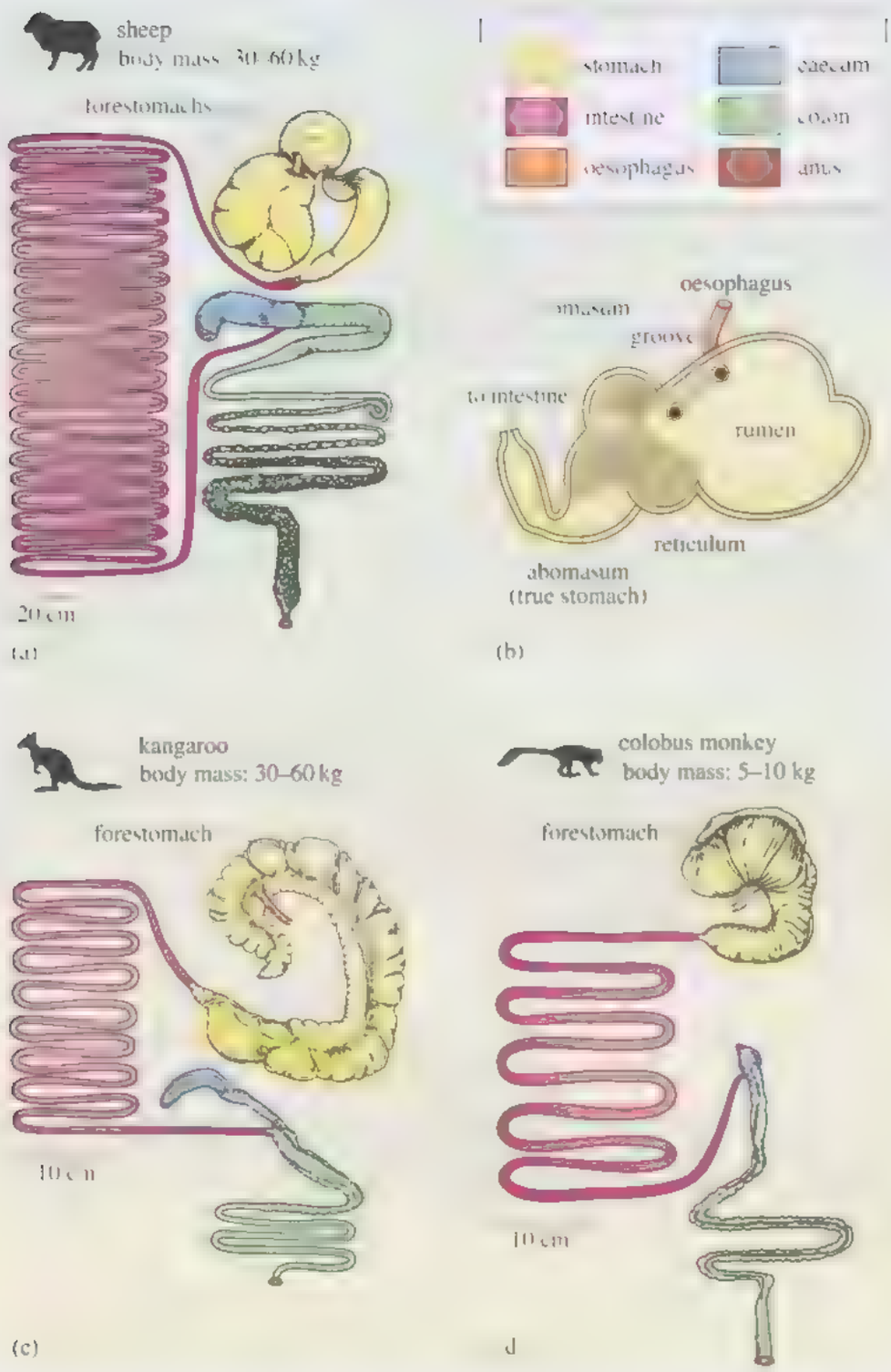



**Figure 2.11** Plans of the guts of some herbivorous mammals that have foregut fermentation. (a) and (b) sheep (a true ruminant), the rumen is huge — in many wild ruminants, larger than all the rest of the gut; (c) kangaroo (*Macropus giganteus*); (d) colobus monkey (*Colobus abyssinicus*).

composition of the fragment. Many microbes can synthesize only some of the enzymes required to complete digestion: those that can digest cell walls forge access for those that preferentially attack starch or proteins. Digestion seems to proceed fastest when several species of bacteria and eukaryotic organisms assemble into structured 'consortia', producing complementary enzymes and often utilizing the products of each other's metabolism.

The microbes produce enzymes that liberate glucose from the cellulose of the plant cell walls, and use it and other nutrients from the plant cells to build their bodies and proliferate. Glucose cannot be oxidized completely because any molecular oxygen associated with the food is quickly used up by the microbes, making the rumen anaerobic. Instead, glucose is fermented to form **short-chain fatty acids**, particularly butyric ($CH_3(CH_2)_2COOH$), propionic ($CH_3CH_2COOH$) and acetic ($CH_3COOH$) acids, sometimes called 'volatile' fatty acids because, being of low molecular mass, they evaporate readily.

These fatty acids are soluble and readily diffuse through the wall of the rumen and into tissues which use them, as well as or instead of glucose or long-chain fatty acids, to produce metabolic energy. Almost all the fuel used by the musculature of the rumen itself is short chain fatty acids. Certain other prokaryotic microbes that live in the rumen can use them too, and some fatty acids may be further broken down to methane ($CH_4$), carbon dioxide ($CO_2$) and hydrogen ($H_2$), which are actively expelled from the rumen as 'burps' through the mouth.

❯ Could the animal oxidize free hydrogen or methane to produce energy?

● No, but the components of hydrogen or methane (H and C atoms) could have provided metabolic energy if they had remained incorporated into glucose or a fatty acid.

These gases are by-products of microbial metabolism that cannot be utilized either by their host or by other symbionts present in the rumen. Amounts thus wasted depend upon diet, but can be substantial. In one investigation, cows eating 5 kg of hay per day released over 190 litres of methane, amounting to 10% of the digestible food intake. Although almost odourless to us, these gases are usually accompanied by small amounts of the volatile fatty acids, which give cattle stables their distinctive smell. If, as happens after the animal's death, microbial fermentation continues but the waste gases cannot be expelled, they accumulate, distending the rumen. Sometimes the gas pressure thus built up is sufficient to burst the abdomen wall.

It might strike you as odd that ruminants harbour microbes that squander valuable plant carbohydrates in this way. The rumen may contain scores, sometimes hundreds of different kinds of microbes, even though the basic digestive processes could be performed by just a few species. The diversity is necessary because the microbes also inactivate or break down plant toxins (see Section 2.3). Often, only a few species of microbes can deal with the toxins produced by each kind of plant that the animal eats. Abrupt changes in forage composition can impair digestion until the appropriate microbes have become abundant enough to handle the new diet. A ruminant cannot control its own digestive capacity as quickly or as directly as non-ruminants can.

By breaking down or neutralizing potentially harmful substances before they reach the small intestine, these microbes prevent the toxins from entering the bloodstream, and relieve the animals' liver of the need to break them down. Resins, tannins and other plant toxins found particularly in mature vegetation and trees are less efficiently broken down. In large quantities, they delay the passage of food through the gut, and impair the absorption of other nutrients.

Thoroughly digested rumen contents flow through the small omasum to the abomasum (see Figure 2.11b), which is homologous to the normal vertebrate stomach. Most of the microbes are killed by contact with the stomach acid and their ruptured cells, together with the remains of the plant tissues and the dissolved products of fermentation, are subject to further digestion before passing into the intestines, where the final digestion products are absorbed as described in Section 2.5.

❯ How would the chemical composition of nutrients absorbed by ruminants' intestines differ from that of nutrients released by digestion of the same material by non-ruminant herbivores?

● The amino acids would come from digestion of the microbes themselves, as well as from the plant protein. There would be very little glucose because most of it has been converted into short-chain fatty acids (or even smaller molecules), which would be far more abundant than after digestion in a simple stomach.

Although ruminants' diet consists mainly of plant polysaccharides, they are actually chronically short of glucose, and sometimes make it from amino acids in the liver. The microbes also alter the plant lipids. In the strictly anaerobic conditions that prevail inside the rumen, many unsaturated fatty acids are saturated by adding hydrogen atoms. These modified lipids are absorbed along with other nutrients and may be broken down for energy production, or incorporated into storage triacylglycerols, or secreted in milk. So meat and milk from ruminants contain a much higher proportion of saturated fatty acids than those of most other mammals.

❯ Can you relate these properties to the relative difficulty of washing up a pan in which mutton or pork has been roasted?

● Mutton (or beef) triacylglycerols, from ruminant sheep (or cattle), contain a larger proportion of saturated fatty acids, which makes them firm, almost hard, at room temperature, while those of pork, from non-ruminant pigs, remain liquid. Hotter water, more detergent and more scrubbing are needed to solubilize the 'hard' fats from ruminant sheep or cattle.

Leaves and stems are low in protein, but ruminants deploy their symbiotic microbes to 'recycle' their own waste products, thereby supplementing the supply of amino acids derived from the diet. As in other vertebrates, the liver converts degradation products of amino acids into urea or ammonia, but instead of being eliminated through the kidney (as happens in most mammals including humans) these 'wastes' go back into the rumen as a component of saliva and by diffusing across the stomach lining. The microbes use these nitrogen-containing molecules to synthesize amino acids, which are incorporated into their own proteins and are then released by digestion and reabsorbed by the animal.

❯ What other advantage would urea recycling confer on ruminants?

● Urea excretion requires water, so urea recycling also reduces urine volume, thus saving water.

Urea recycling is particularly efficient in camels, which thus thrive on low-quality forage containing very little protein, and excrete only small amounts of urine.

As well as re-using the nitrogen of degraded proteins, symbiotic microbes also synthesize most of the water-soluble vitamins, especially those of the B group. Animals absorb them from the digested remains of the microbes, and not just the herbivore that was carrying the symbionts in its stomach: most predators eat their prey's guts, and at least a fair amount of their contents. Distasteful though it may seem, this habit makes an essential contribution to nutrition: captive lions and tigers fed only on eviscerated carcases eventually develop vitamin deficiencies.

Several other lineages of mammals, including hippopotamuses, kangaroos (Figure 2.11c), and several kinds of monkeys (colobus, langur and proboscis monkey, Figure 2.11d) have an enlarged, non-secretory forestomach in which microbes ferment plant material before it reaches the acidic portion of the stomach. In this respect, they resemble ruminants, but they do not chew the cud, and certain kinds of microbes, notably ciliates (protoctists), are usually absent.

How would the absence of cud-chewing affect forestomach fermentation?

- The food is less finely fragmented, prolonging digestion time, so the forestomach has to be very large relative to the size of the body.

Hippopotamuses do not chew cud, but the first of their three stomachs is huge: its contents alone have been measured as 220 kg after a night's grazing, or 13–15% of the body mass. Hippos spend most of the day resting in pools and rivers, thus taking the enormous weight off their legs. The end-products of this prolonged digestion are smooth, viscous faeces that can be broadcast by flicking the short, bristly tail, as mature males do to establish breeding territories.

### 2.4.4 HINDGUT FERMENTATION

In various other herbivorous mammals, including all perissodactyls (horses, zebras, tapirs and rhinoceroses), rodents (rats, guinea pigs and squirrels), lagomorphs (rabbits and hares), elephants and some metatherians such as the koala bear, symbiotic microbes are concentrated in an enlarged colon and/or caecum. The gut contents thus reach the microbes after the animals' own enzymes have digested them as much as possible. Figure 2.12 shows the proportions of the stomach, caecum, colon and other components of the gut in various mammals.

What would be the disadvantages of this arrangement compared to forestomach fermentation?

- The products of microbial digestion of cellulose cannot be absorbed unless the material re-enters the small intestine, and the bacteria themselves cannot be digested.

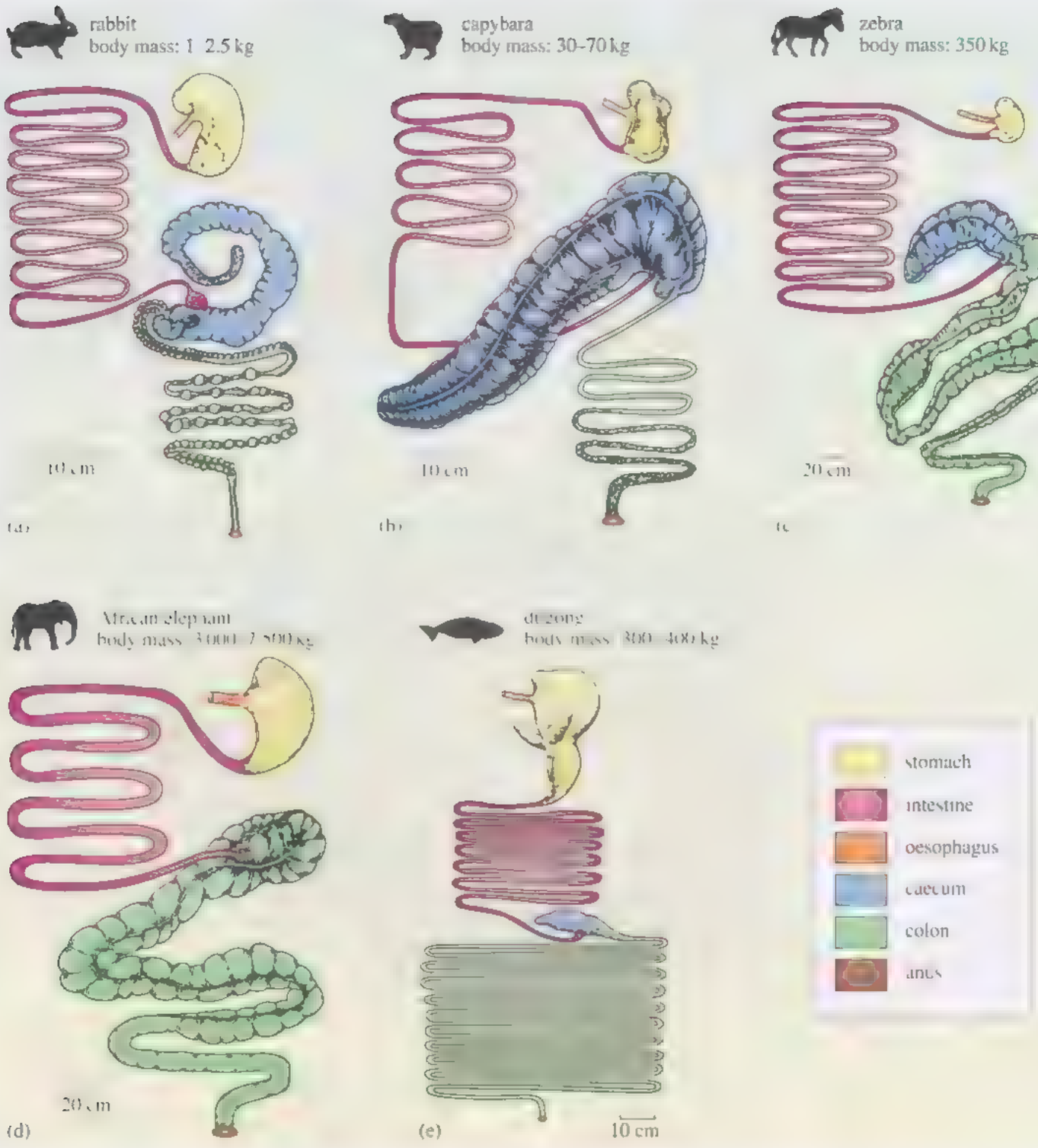



**Figure 2.12** Plans of the guts of some herbivorous mammals in which microbial fermentation takes place in the colon or caecum: (a) rabbit (*Oryctolagus cuniculus*); (b) capybara (*Hydrochoerus hydrochaeris*), a large rodent similar to guinea-pigs; (c) zebra (*Equus burchelli*); (d) African elephant (*Loxodonta africana*); (e) dugong (*Dugong dugong*).

An obvious solution to this problem is intestinal reflux: in horses, material leaving the caecum probably moves backwards into the absorptive area of the small intestine before it goes to the colon and is expelled. Lagomorphs and some rodents, notably coypus, achieve much the same effect by eating their faeces, a habit known as **coprophagy**. The 'first pass' faecal pellets are soft and fibrous, and rabbits eat them as soon as they emerge, usually in the privacy of the burrow. The dark-coloured, homogenous rabbit droppings often seen on grassland are the product of the second passage through the gut.

Recent studies of wild mammals reveal that coprophagy is more widespread than was previously believed. Svalbard reindeer, *Rangifer tarandus platyrhynchus*, live and breed as far north as 80 °N, where the vegetation is very sparse and is covered with snow for nine months of the year (Figure 2.13). They supplement their meagre, indigestible diet by eating the fresh faeces of barnacle geese, *Branta leucopsis*, that migrate to breed on Svalbard during the brief summer. Experiments show that the reindeer choose goose faeces in preference to vegetation, and such 'reprocessing' of incompletely digested plants makes a significant contribution to their nutrition.

**Figure 2.13** *Rangifer tarandus platyrhynchus*, grazing on Svalbard in late May.

What do these habits suggest about the efficiency of plant digestion in geese and reindeer?

- Geese do not digest their food as thoroughly as reindeer can, but passage through their gut facilitates further digestion in the reindeer's rumen.

Koala bears eat eucalyptus leaves, which are almost as tough and indigestible as the Svalbard reindeer's diet. The single offspring is suckled for an unusually long period, and is slowly weaned onto its mother's faeces, which both supply nutrients and colonize the young's gut with appropriate symbiotic microbes, before it tackles the adult diet.

Hindgut fermentation might seem to be a poor substitute for digestion in a ruminant forestomach, but careful observations of the composition of food ingested and faeces eliminated, and of the ecological circumstances under which animals with

different kinds of digestive system prosper have revealed some advantages. For example, food passes through the gut more quickly in hindgut fermenters, taking 30–45 h in a horse compared to 70–100 h in a cow of similar body mass. Faster transit time is important if a large proportion of the forage is totally indigestible.

Silica, the component of grasses that makes their leaves and stems sharp and abrasive, is chemically similar to sand and is not digested at all. Resins and tannins are organic compounds that in principle could be converted into metabolically usable substances. Some beetles, caterpillars and other herbivorous insects can utilize them, but vertebrate herbivores cannot. These compounds simply occupy space, and in a forestomach they clog up the flow, thereby slowing down digestion and absorption of useful materials. The faster passage through the gut of hindgut fermenters enables this indigestible material to be expelled more promptly. Fragments of straw and other tough plant parts are abundant and clearly visible in the droppings of horses, rhinoceroses and elephants, but cannot be seen in the faeces of cows, sheep or deer.

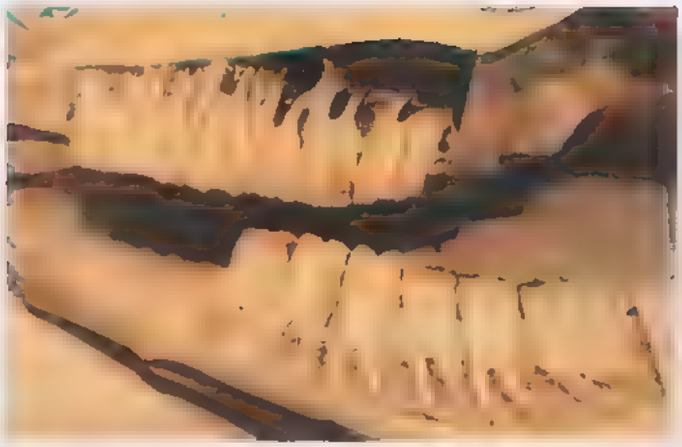
Because hindgut fermenters extract energy from plants less efficiently, they need more forage to sustain themselves and produce much more faeces than carnivores or ruminant herbivores. African elephants (Figure 2.12d) defaecate 14–20 times per day, producing a total of 150 kg wet faeces containing 35 kg dry matter. This quantity is produced from about 60 kg of dry matter eaten, indicating that only about 40% is digested well enough to be absorbed. Gathering this amount of food takes around three-quarters of the elephants' time, leaving only a few hours for rest and sleep. The longest colon in Figure 2.12 is that of the dugong (2.12e), a distant relative of elephants which lives permanently in water, grazing aquatic plants in estuaries and shallow coastal waters. Weight in water is always much less than weight in air, so aquatic animals can afford to have a long, heavy gut in which large quantities of forage are digested slowly but thoroughly.

Zebras (Figure 2.12c) also have long intestines and survive seasonal droughts by eating large quantities of dry, tough plant material and digesting it incompletely but quite fast.

❯ How would this diet and digestive physiology affect the teeth?

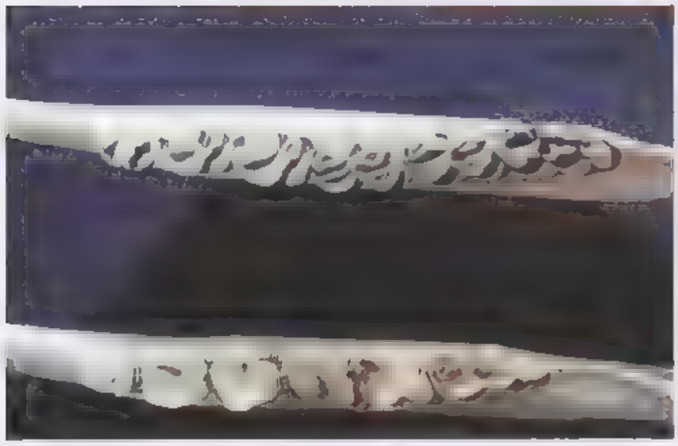
● The teeth would wear down faster because hindgut fermenters eat more food, and more of it is likely to be very abrasive grasses.

The teeth of equids (horses, zebras and donkeys) withstand such use because the roots remain open, and the very long crown erupts slowly but continuously throughout life, thus replacing surface that is worn away (Figure 2.14a). Elephants combat tooth wear in another way: each tooth is huge, the largest weighing several kilograms. The three premolars of the 'milk' set and the three molar teeth in each jaw are used in succession instead of simultaneously, each tooth moves forward in the jaw as it is worn in use, to be replaced from behind by the next, larger tooth (Figure 2.14b). In these herbivores, which eat tough, abrasive plants, the dentition sets the maximum lifespan: old elephants, reindeer (Figure 2.14c) and donkeys die from starvation when their teeth wear down, so they cannot chew their food into particles fine enough for digestion to be efficient.



(a)

(b)



(c)

**Figure 2.14** (a) Section of a horse skull to show long, continuously erupting teeth. The bone has been cut away from both the upper and lower jaws to reveal the complex roots of the molar teeth. The wear surface is visible on the lower jaw. (b) The lower jaws of a youngish African elephant. At the time of death, only a single molar tooth in each half of the jaw was in use as a grinding surface. Another tooth is in the process of formation behind each of the functional molars, but in life they were probably not yet erupted through the gum. (c) The molar teeth in the lower jaw of a young (top) and elderly (below) wild Svalbard reindeer.

The need to eat large quantities of plants helps to explain why zebras feature so prominently in safari tours: the buses usually pass around midday when most of the ruminants have retired to shady glades to chew the cud from their early morning foraging, before emerging to feed again in the cool of the evening. Especially during the dry season, when food quality is poor but tourists abundant, zebras eat almost continuously, and often do so in full sunshine, where it would be too hot for the ruminants. They are thus more likely to have their photos taken than the ruminants, which are certainly more diverse and, in most places, more numerous.

How would the differences in digestive physiology affect the composition of fuels circulating in the blood?

- Glucose liberated from sugars and starches digested by the animals' own enzymes is absorbed in the small intestine before the microbes get a chance to convert it to short-chain fatty acids, so the muscles and other energy-utilizing tissues of hindgut fermenters use far more glucose and less fatty acids than do those of ruminants.

Animals' digestive physiology determines many other aspects of their metabolism. Like other non-ruminants, horses use mainly glucose to fuel brief bursts of strenuous exercise, and, like human athletes, their performance in

activities such as short distance racing can be enhanced by feeding them on easily digested, starch-rich foods such as oats. The extra glucose is stored in the liver and the muscles as glycogen, which can be quickly broken down into glucose again, providing a ready supply of fuel for the muscles. The absorption of plenty of glucose also means that mares' milk contains much more sugar (and less fat) than that of cows and other ruminants. Pastoralists who herd horses on the steppes of central Asia have long fermented mares' milk to form an alcoholic drink called koumiss. Bacteria convert the sugars to alcohol in the same way as the carbohydrates in grapes, grain, potatoes, bananas or sugar cane can be fermented. Ruminant milk is much less effective for this purpose because it contains too little sugar and too much fat, so the end product of bacterial action is yoghurt or cheese.

### SUMMARY OF SECTION 2.4

1 Digestion of organic material involves mechanical shredding and churning, and enzymatic breakdown of large molecules.

2 The stomach and intestine, and associated glands, secrete digestive enzymes. Some are highly specific and attack only certain kinds of bonds in certain positions within substrate molecules.

3 Bile salts secreted from the liver emulsify lipids, thereby facilitating their digestion by lipases. The capacity for digesting waxes is low in most birds, mammals and insects, but can be much higher in animals whose diet includes large quantities of wax.

4 Efficient digestion depends upon adjusting the concentrations of food and enzymes. Water secreted and mixed with food in the mouth or stomach is reabsorbed in the hindgut. Very dilute food may be useless to animals that cannot excrete the excess water quickly and efficiently.

5 Most vertebrate herbivores depend upon symbiotic microbes to digest tough or toxic plants. A forestomach or modifications of the hindgut that accommodate microbes have evolved several times among mammals, but only ruminants chew the cud.

6 Microbial fermentation in the rumen alters energy metabolism in all the other tissues and improves protein nutrition by recycling of urea and/or synthesizing essential amino acids.

7 Intestinal reflux and/or coprophagy allow absorption of the products of microbial digestion in hindgut fermenters, thus improving nutrient extraction from plant food.

## 2.5 ABSORPTION AND METABOLISM

The small molecules generated by digestion of food are absorbed through special epithelial cells lining the small intestine, into the blood or other body fluid to reach tissues elsewhere. Carbohydrates are absorbed as monosaccharides (e.g. glucose or fructose), proteins as single amino acids or small peptides, and lipids as fatty acids or monoacylglycerols. Small quantities of monosaccharides and peptides pass through the gut epithelium by diffusion, but the majority, at least in mammals, are taken up molecule by molecule by **transporters** (or **carriers**).

assemblages of specific proteins situated on and within the membrane of the epithelial cells on the side facing the gut contents (see Figure 2.1). Some such transporters are passive, in that they function without using energy, but **active transport** requires metabolic energy in the form of ATP and can take up substances against their concentration gradient, i.e. moving them from a more dilute solution into a more concentrated one.

In most vertebrates that have been studied in detail, the intestine has a special transporter for glucose and galactose, another for fructose, and separate transporters for each of the major categories of amino acids. The overall rate of absorption of any particular amino acid or sugar is determined by the activity of the appropriate type of transporter, their number per unit area of gut lining and the area of absorptive epithelium that is exposed to the gut contents. Comparisons between species suggest that the latter is the most important. Mammals absorb nutrients 7–10 times faster than lizards of similar body size, primarily because they have more villi, which make their intestinal surface area 4–4.5 times greater than that of reptiles. Although the mammalian body temperature is continuously higher, and they are generally more active so need much more energy, the maximum rate of transport per unit gut area is only about twice that of reptiles.

## 2.5.1 ADAPTING ACTIVE TRANSPORT TO DIFFERENT DIETS

Modern methods of protein separation and identification allow specific transporters to be studied individually, but from a functional point of view, it is often more useful to measure the total uptake capacity of the intestine. A long-established technique is to turn the small intestine of a freshly killed animal inside out, tie both ends and immerse it for an hour or two in a solution containing amino acids or monosaccharides, then measure their accumulation in the closed tube formed by the inside-out intestine. Figure 2.15 shows such data for glucose and proline, a common component of proteins. The tissues came from laboratory mice that had been fed for up to 2 weeks on synthetic diets containing different proportions of protein and carbohydrate.

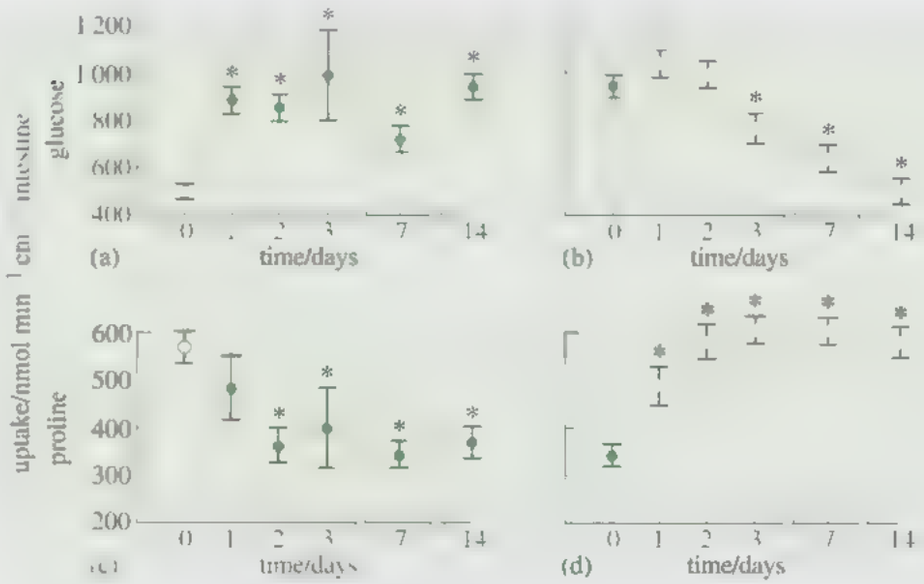



**Figure 2.15** Adaptive changes in uptake of glucose (graphs (a) and (b)) and proline (graphs (c) and (d)) in mice fed on high-protein food without carbohydrate (open circles) then switched to a high-carbohydrate, medium-protein rations (solid circles) for two weeks (graphs (a) and (c)), then back to high-protein, carbohydrate-free food (graphs (b) and (d)). In each case, the dietary change took place on day zero. Samples of mice were killed each day and the capacity of their small intestines to take up glucose or proline measured. * indicates significant difference from the values measured on day 0. Data from Diamond and Buddington (1987).

From the data in Figure 2.15, what happens when the proportion of carbohydrate in the diet is abruptly changed?

When dietary carbohydrate is increased, uptake of glucose rises by at least 50% within a day (Figure 2.15a), but the reverse adjustment is slower, taking a week to complete (Figure 2.15b).

The effects of the experimental diets on the uptake of proline were more symmetrical, being almost complete within 2–3 days of the change in diet (Figure 2.15c and d).

What general conclusion can you draw from the data in Figure 2.15?

When the diet and/or digestive mechanisms yield little glucose or proline in the small intestine, the appropriate transport mechanisms in its epithelium decline, and the converse.

Although reduced, the transporter systems do not disappear completely: some uptake capacity remains, even when there is apparently little for it to do.

Would nutrient uptake be efficient for mice abruptly switched from one diet to the other?

No. For the first few days, the new type of food would not be absorbed (and probably not even digested) properly.

Omnivorous animals, which normally eat a wide range of different foods, maintain broader digestive and absorptive capacities. Figure 2.16 shows some data obtained by feeding the same synthetic diet to several species of carnivorous, omnivorous and herbivorous fish.

Are the short-term adaptations to high-protein diet experimentally induced in mice (Figure 2.15) as large as the natural differences between fish specialized to carnivory or herbivory (Figure 2.16)?

No. The ratio of proline uptake to glucose uptake reaches 5–6 in the carnivorous fish, but the maximum ratio induced by feeding the mice on a protein enriched diet was about 600/500 = 1.2.

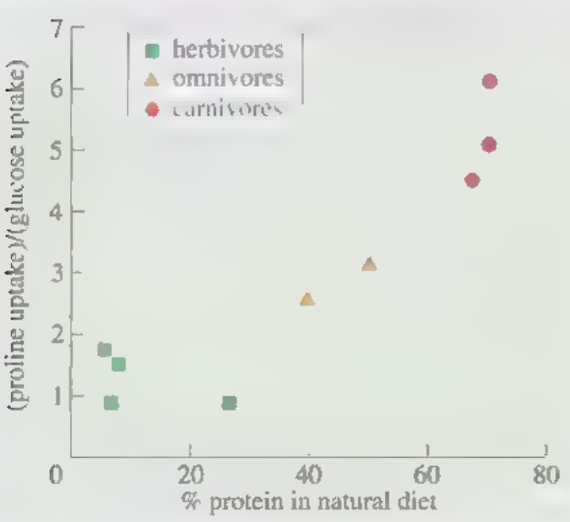



**Figure 2.16** Capacity for absorbing glucose and amino acids (estimated as the ratio of uptake of proline to uptake of glucose) in the intestine of various species of carnivorous, omnivorous or herbivorous fish. At the time of the experiment, all the fish were kept in the lab and were fed the same synthetic food. The proportions of protein in the species' natural diets differed, as indicated on the horizontal axis. Data from Diamond and Buddington (1987).

As in so many physiological systems, inheritance defines the limits to physiological capacities, but within that range, actual properties can be adjusted to immediate requirements. In evolutionary biology parlance, there is a 'trade-off' between economical, thorough digestion, and the capacity for opportunistic exploitation of a wide variety of foods. Rats, mice and people are dietary generalists, but many insects, including many kinds of aphids, beetles and caterpillars, naturally feed on only one plant species and they can digest, absorb and utilize components of that diet that are useless, even toxic, to all other animals (see Chapter 5). Their food selection behaviour is so firmly tied to digestive ability that they may refuse to eat anything other than their own host plant.

#### GUTLESS FEEDING

The capacity for active absorption of nutrients is not limited to the cells lining the gut, and is not necessarily preceded by the secretion of digestive enzymes. The gut may be reduced or absent in animals that are surrounded by nutritive materials in solution, and substances are absorbed at the body surface, as in osmotrophic protoctists.

) What kinds of animals would abandon digestion in this way?

● Internal parasites such as tapeworms, which live inside the intestines of other animals.

Their hosts secrete digestive enzymes, so the parasites do not do so, but they retain the capacity for active uptake of the solubilized nutrients, in many cases in competition with the host itself.

Some free-living animals, including certain marine corals, worms and clams, can take up dissolved amino acids and glucose, and possibly other nutrients, at much lower concentrations, from a few milligrams to micrograms per litre. This style of feeding may require a lot of energy for active transport, but its total energetic cost might still be low compared to that required to obtain the same nutrition by filtering particles from water or chasing after prey, then deploying digestive enzymes to break it down.

### 2.5.2 MINERAL ABSORPTION

Transporter systems are known for many minor nutrients, including certain vitamins, and minerals such as iron, calcium, zinc and copper, at least in birds and mammals. It has been possible to demonstrate that decreasing the amount of certain vitamins or minerals in the diet leads to increases in their rates of transport through the gut epithelium, and conversely, feeding excess of the nutrient reduces its uptake. But the abundance of transporters is often very low, in many cases, too low to measure accurately.

) What kinds of diets and feeding habits are likely to present the least problems with mineral nutrition?

● Carnivory, especially if predator and prey are of similar chemical composition (e.g. snakes eating other vertebrates, spiders eating insects) and the prey is eaten whole.

Obtaining enough minerals is a major problem for all herbivores because most plant tissues contain very small amounts. Sodium ions are essential to all animals because of their role in the transmission of signals between cells, especially those of the nervous system. In contrast, plants do not seem to need sodium at all, so most contain almost none, and only a few species (e.g. salt-marsh and beach plants) can tolerate soils that contain more than small quantities. Calcium and magnesium are important for many intracellular processes in both plants and animals, but vertebrates need them in much larger quantities to support the growth of bones and teeth, as do shelled molluscs. Herbivorous animals have various means of supplementing diets that lack these essential minerals.

The aquatic larvae of mosquitoes feed on detritus and microbes, but the adult females must have at least one meal of vertebrate blood for their eggs to mature. Blood is believed to provide iron, sodium and other minerals that are lacking in the larval diet, as well as being a rich source of protein.

Elephants actively stir up stream beds to drink the fine mud so formed, and eat the mineral-rich bark or sapwood of certain trees, especially the baobab or bottle tree (Figure 2.17). Buffalo and other ruminants travel long distances to eat soft rocks that are rich in sodium or calcium, and they may lick the blood of injured animals, including each other. These large mammalian herbivores and smaller species such as rats, voles, rabbits, and even large herbivorous birds such as parrots, sometimes chew dried bones, especially while breeding. Smaller birds take advantage of spring rains to catch earthworms, ingesting the earth in the gut as well as worm flesh.

**Figure 2.17** A baobab tree in Kenya damaged by elephants ramming their tusks into the trunk to eat the mineral-rich bark and sapwood

Omnivorous and carnivorous mammals eat their own placentas, and may eat their new-born young if seriously disturbed while suckling. Most 'herbivorous' mammals, including mice and rabbits, do the same, showing that they can manipulate and digest meat: they do not normally do so because they can deal with other foods more efficiently. However, ruminant mothers leave whole placentas and dead calves and fawns for vultures and other scavengers. The rumen microbes cannot deal with large chunks of meat and might be poisoned by it.

The shells of terrestrial and freshwater molluscs are rarely as massive as those of marine species (seawater contains plenty of calcium) and many species are restricted to chalky soils, which are rich in calcium.

Active transport is usually presented as a means of scavenging for scarce substances, as in the case of rare vitamins and minerals. But it can also be a means of limiting uptake, thus avoiding taking in too much of potentially harmful substances.

❯ Which two major properties of active transport enable it to act in this way?

● (1) Active transport can be regulated according to supply and demand

(2) There are separate transporter systems for each kind of nutrient (or small group of very similar nutrients) so uptake of one can be adjusted without affecting the others.

Although essential in small quantities, many vitamins, especially the fat-soluble vitamins A and D, are toxic when present in excess. The same is true of many of the essential minerals, including iron, a highly reactive transition element. Vertebrate blood is unusually rich in iron because a large proportion of the red cells' contents is the oxygen-binding protein, haemoglobin, which contains iron atoms. Blood-sucking animals such as leeches (phylum Annelida, class Hirudinea), ticks (phylum Arthropoda, class Arachnida), bedbugs (phylum Arthropoda, class Insecta) and vampire bats (class Mammalia, order Chiroptera) can digest the protein component of haemoglobin and absorb the amino acids so released (together with the glucose, lipids and other nutritious components of the blood), but most of the iron itself remains in the gut and is expelled in the faeces.

› What kinds of organisms would be unable to deal with too much iron in this way?

● Unicellular organisms, which digest all their food intracellularly.

The protoctist *Plasmodium* spends part of its complicated life cycle in the blood of mammals and birds, where it feeds by boring into red blood cells and consuming their contents. The parasite proliferates until the host's cell bursts, and the progeny escape to infect more cells. *Plasmodium*'s food vacuoles secrete not only the usual proteases, which digest most of the proteins, but also a special enzyme that builds the iron-containing proteins into insoluble, and thus non-toxic, crystals. The parasites are so small, and chemically so similar to the hosts' cells, that they are very difficult to see with a light microscope. The dark, iron-containing crystals were the distinctive feature that enabled the French army surgeon, Alphonse Laveran, to recognize them, and to suggest, in 1880, that they were associated with malaria.

Recent research has shown that drugs derived from quinine, the ancient preventative and cure for this widespread disease, work by inhibiting the enzyme that promotes crystallization of the iron-containing parts of haemoglobin. When the drug is present, the iron is released by digestion and diffuses into the rest of the parasitic cells, poisoning them.

› Why would the drug not harm the cells of the parasite's host?

● They do not need, and do not have, the crystal-forming enzyme because almost all of their iron is locked up inside protein molecules.

The free iron that leaks from dying *Plasmodium* never becomes concentrated enough to harm the host's cells.

Unfortunately, several strains of the *Plasmodium* species that cause malaria in humans and other primates have recently developed physiological means — we do not know exactly what — of preventing quinine-like drugs from inhibiting this unique and crucial enzyme. They have thus become 'drug-resistant', and can only be suppressed effectively by much larger doses, which can harm their hosts by other mechanisms.

### 2.5.3 NON-ACTIVE ABSORPTION

The absorption of lipids does not involve specific transporters that utilize metabolic energy. Although most fatty acids and monoacylglycerols are larger molecules than single amino acids or glucose, they diffuse through cell membranes because they consist mainly of lipids. Until recently, all products of digestion of triacylglycerols were believed to diffuse unassisted into the gut epithelium, but during the 1990s, an assortment of carrier molecules and binding proteins were identified that facilitate the passage of long-chain fatty acids through cell membranes without utilizing metabolic energy.

Although lipids readily enter cells, conveying them between cells is more difficult.

> How would features of glucose and amino acids that contrast with those of lipids affect their mobility in aqueous solutions such as blood?

- Glucose and amino acids would dissolve readily, but long-chain fatty acids and triacylglycerols would be as insoluble in the blood as they were in the gut contents.

Glucose and amino acids (and most vitamins and minerals) are in free solution in the blood, but special molecular apparatus is needed to escort lipids around the body. The long-chain fatty acids and monoacylglycerols are re-esterified into triacylglycerols in the absorptive epithelium and assembled, with a few structural proteins, into **chylomicrons** before being released into the blood (and, in the case of mammals, the lymph ducts). Chylomicrons can be formed very quickly and in large numbers, and may become huge by biochemical standards. The large ones are up to 0.001 mm in diameter (they average around 70 nm ($7 \times 10^{-8}$ m)), too big to squeeze into the space between adjacent cells of most tissues, and only a little smaller than the bore of small blood vessels.

After a fatty meal, chylomicrons may be big enough and numerous enough to produce a noticeable change in the colour and fluidity of the blood, but they only last a few minutes: adipose tissue (and some other tissues, such as muscle) synthesizes lipoprotein lipase and releases it into the tiny blood vessels that permeate the tissue. This enzyme severs the fatty acids from the triacylglycerols, fragmenting the chylomicrons.

The gut lining is never completely impermeable, and a wide range of substances can pass, at least in small quantities, from the gut to the blood with no physiological assistance at all. Most small, water-soluble molecules are quickly excreted if they are non-nutritive, and many potentially toxic organic materials are degraded (by the liver in vertebrates, or by analogous tissues in invertebrates) to harmless products that can be excreted. But heavy-metal ions such as mercury and lead, and synthetic organic compounds such as DDT and PCBs cannot be excreted, so they gradually accumulate in the body until they reach concentrations that hinder vital functions.

Damage to the nervous system, sensory organs or liver usually produces the most noticeable symptoms of such poisoning, but some absorbed pollutants have more subtle effects, on gamete formation or on the skeleton. Passive absorption of toxins in food or drinking water is becoming more of a hazard now than in the past, because mining and smelting release more heavy-metal ions into water and soil, and a much wider variety of completely new substances are being synthesized, a few of which are absorbed surprisingly well. For example, soluble derivatives of certain plastics permeate animal guts and/or gills and interfere with the hormones that control sexual maturation and sperm formation. They may accumulate in rivers and ponds, where they impair the breeding of amphibians and fish, and possibly also of people who drink the water.

### 2.5.4 HERBIVOROUS INSECTS

As mentioned in Section 2.5.2, many plants lack adequate quantities of nutrients that are essential to animals, especially large, active or fast growing species.

❯ How do vertebrates and molluscs avert such nutritional problems?

● By eating a wide range of different species, and by supplementing their diet with minerals and small quantities of animal material (see Sections 2.4.3 and 2.4.4).

Such strategies are not available to many insects which are highly specialized to feed only on one component (leaves, flowers, roots, etc.) of one or a very few species of plants. One of the most meagre and monotonous diets is that of aphids: throughout their lives, they feed only by sucking the sap of green plants through highly specialized mouthparts that pierce the outer layers of stems to reach the sugar-transporting tissue (phloem) beneath, as shown in Figure 2.18.

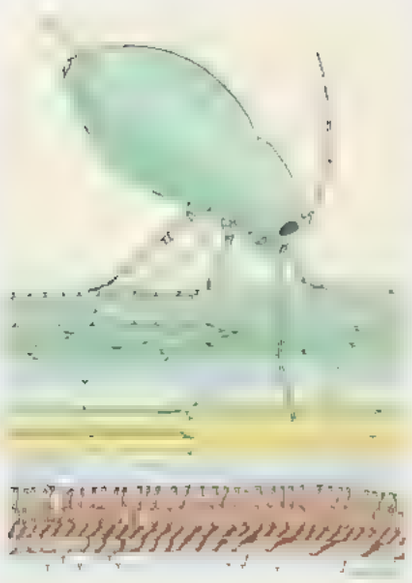


**Figure 2.18** An aphid sucking from a plant stem.

Phloem sap is a dilute solution of sucrose and water, plus small quantities of amino acids, mostly the highly soluble amino acid glutamine, and a few minerals, mostly potassium ions.

❯ Could rats, cows, bumblebees or people survive on such a diet?

● No. Sap lacks the wide range of minerals, amino acids and fatty acids and proteins that are essential to vertebrate herbivores for synthesizing their own body fabric.

Nonetheless, aphids thrive and proliferate amazingly fast. Exceptionally for insects, most species are viviparous, i.e. they give birth to well-developed young rather than lay eggs (see Figure 2.19a). For example, the greenbug, *Schizaphis graminum*, weighs about 24 µg at birth, grows to 540 µg in 10 days, and can breed at 11 days old. Aphids can have up to 15 generations in a summer and the progeny of a single female may form dense assemblages covering rose buds, pea and bean flowers and many other soft plant tissues. The insects have a voracious appetite for sap, often taking in many times their own body mass each day and making their host plant wilt. However, much of the sugar and water is quickly excreted through the anus as 'honey-dew'.

❯ Why is excretion of sugar unusual?

● Most animals excrete little or no sugar because they can utilize glucose as fuel for movement and biosynthesis.

Glucose fuels the aphids' rapid growth and is incorporated into chitin, the polysaccharide that is a major component of their cuticle, but they hardly move about at all except during the winged, reproductive phase which appears towards the end of the summer. To obtain enough of other nutrients, aphids take in more sugar and water than they can use, so the excesses are eliminated, sometimes in such large quantities that on warm days, aphid-infested trees appear to drip syrup. This honey-dew is nutritious to physically active animals, and many aphids are attended by ants that 'milk' their droppings. Certain aphids that feed on fir trees are attractive to bees which collect honey-dew as well as or instead of nectar. In the Black Forest in Germany, such bees are managed for the production of 'fir honey', one of the most highly valued kinds of honey.

❯ Could the ability to excrete honey-dew correct the deficiencies of protein, lipids and minerals in their diet?

● No. Elimination of excess sugar enables the insects to process larger volumes of sap from which the small quantities of glutamine and other dilute nutrients could be extracted. But the habit would not solve the basic problem of the lack of essential nutrients in the diet.

❯ What kinds of organisms can synthesize a wide range of amino acids and lipids and certain vitamins from a few simple precursors?

● Microbes such as bacteria perform this role in the forestomach of ruminants (and elsewhere).

During the 1990s, the combination of high-powered electron microscopy, molecular characterization of gene structure and function, and comparative studies of a wide range of different aphid species has revealed that these insects owe their ability to thrive on an apparently inadequate diet to symbiosis with bacteria of the genus *Buchnera*. These endosymbionts live inside vesicles called symbiosomes (Figure 2.19b), hundreds of which stuff special cells called bacteriocytes (Figure 2.19c) which form a bilobed structure inside the aphid's large abdomen.

The bacteria can synthesize at least four of the amino acids that the aphids need for growth, from the glutamine obtained from the plant sap, and may also contribute vitamins. *Buchnera* enters the embryos early in development, and proliferates as their host grows. Adult aphids weighing 500 µg may contain $5.6 \times 10^6$ *Buchnera* but numbers decline sharply as breeding ends and few are left when the aphids die at the age of about 25 days.

❯ How could viviparity facilitate the transmission of the endosymbionts between generations?

● The developing embryos share the aphid's abdomen with the bacteriocytes

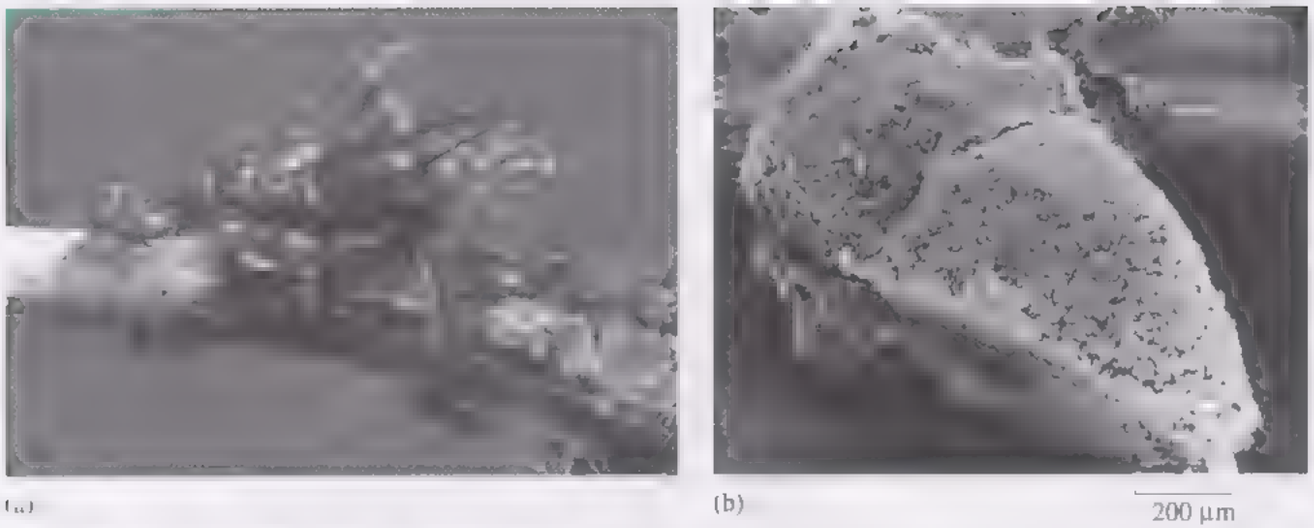



**Figure 2.19** (a) Adult and juveniles of the aphid *Schizaphis graminum*. The large female in the centre is giving birth. The juveniles attach themselves to the stem beside the mother. (b) Scanning electron micrograph of a bacteriocyte stripped of its cell membrane to reveal hundreds of symbiosomes containing *Buchnera*. (c) Transmission electron micrograph of part of a bacteriocyte to show cross-sections of *Buchnera in situ*.

This symbiotic relationship is ancient, dating back at least 200–250 Ma, and the two organisms have evolved together, so that the great majority of the more than 3000 living species of aphids harbour *Buchnera*. *Buchnera* is not found outside aphid cells, though presumably its remote ancestors were once free-living, and as yet, they cannot be maintained in artificial cultures.

Microbial symbionts also contribute to digestion and/or nutrition in a wide range of other herbivorous insects, including other hemipterans such as whiteflies, mealybugs, and cicadas, certain cockroaches, ants and beetles. Such relationships with microbes may have allowed these insects to evolve highly specialized diets and feeding mechanisms, which in turn promoted the formation of a huge range of species. Among the most economically important are wood-eating termites which harbour symbiotic flagellates, sometimes in numbers amounting to 30% of their total body mass, in a special pouch off the hindgut. Termite colonies containing millions of insects thrive on little more than wood and roots. As well as hollowing out large trees, termites destroy building and fencing timber in tropical countries and have recently spread into southern Britain as our climate becomes warmer.

### SUMMARY OF SECTION 2.5

1 Absorption involves specialized cells which in vertebrates are found in the small intestine. The absorptive surface area is greatly increased by being frilled out to form villi.

2 In most animals, a large fraction of the amino acids and sugars released by digestion and many vitamins and minerals are taken up by specific transporters, some of which utilize ATP. They are located in the membrane of the epithelial cells lining the small intestine, but similar transporters may occur on the outer body surface of animals without guts. Fatty acids and monoacylglycerols enter cells by diffusion, often facilitated by carrier proteins.

3 In mice, the capacity for active transport takes at least 1 day to adapt to large changes in diet composition.

4 Many herbivores compensate for the nutritional insufficiencies of plants by eating mineral-rich soil or small quantities of animal food.

5 Unicellular protoctists that cannot avoid taking up too much iron sequester it in an insoluble form.

6 The intestine is not impermeable, and many other substances, including potential toxins, can cross its epithelium and enter the bloodstream.

7 Glucose, amino acids and short-chain fatty acids dissolve in blood but triacylglycerols form chylomicrons.

8 Most aphids harbour one species of endosymbiotic bacteria which synthesize essential amino acids and enable the insects to breed rapidly on a nutritionally meagre diet of sap. Other insects have gut symbionts that aid digestion as well as metabolism.

## 2.6 FEEDING OR FASTING?

Digesting and absorbing food, or even just keeping the gut fully functional, require considerable amounts of energy and materials. Although the gut epithelium (see Figure 2.1) is partially protected from digestion by its own enzymes, mechanical friction as well as chemical attack quickly wear the delicate cells, which are constantly being renewed. Some of the debris so formed is digested and its constituents absorbed, but much is eliminated with the faeces, representing a significant loss of body fabric. Mammalian faeces contain enough sloughed off cells for the individual that produced them to be identified using DNA fingerprinting techniques.

Replacement of the 'front-line' epithelial cells that secrete enzymes and absorb nutrients is also essential for adaption to changes in diet. In mammals, absorptive cells form in the 'crypt' at the base of the villi (Figure 2.1b) by division of small, unspecialized 'stem cells'. Newly formed cells enlarge and develop fully functional transporters, and migrate towards the tip, where they are eventually sloughed off. Cells from different points between the crypt and the tip of the villi can be separated and the number of glucose transporters on them estimated by measuring the binding of radioactively labelled phlorizin, a sugar-like molecule

with a much higher affinity for the transport sites than glucose itself. Figure 2.20 shows the results of such a procedure from the intestinal cells of mice that had been abruptly switched between high and low carbohydrate diets.

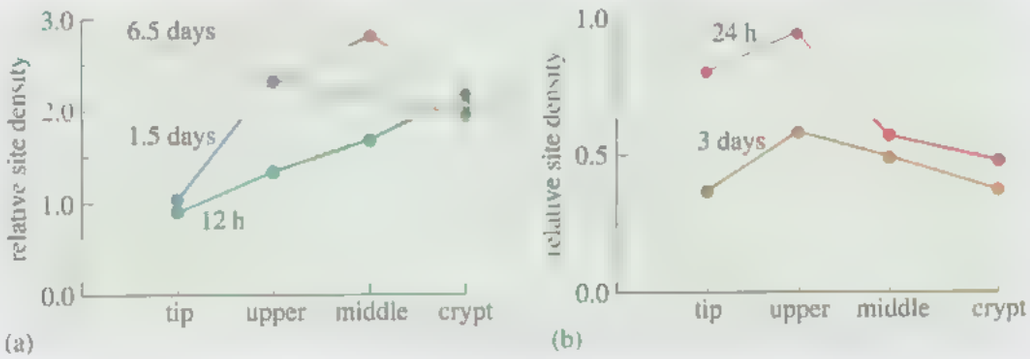



**Figure 2.20** Ratio of the density of phlorizin binding to epithelial cells from four sites along mouse villi measured at various times after a change in diet, to that measured just before the dietary change: (a) change from no digestible carbohydrate to food containing 55% sucrose (a readily digestible carbohydrate); (b) change from a high-carbohydrate diet to food containing 70% protein, 7% fat and no digestible carbohydrate.

❯ When do the cells start to adapt to the change in diet?

● Within 24 h. At the 12 and 24 h sample times (Figure 2.20a and b, respectively), cells in the middle of the villus and in the crypt have acquired binding capacity appropriate to the new diet, while the older cells located towards the tip are adapted to handle the old diet (i.e. relative site density = 1).

The data for the 3-day and 6.5-day samples show that, by then, all the cells are fully adapted to the new diet. The rate of formation of new cells exactly matched the timing of the decline in phlorizin binding, indicating that adaptation of the whole system is due entirely to cell replacement. The old cells cannot change their properties; they are simply replaced by new ones that matured with different abilities.

❯ Would you expect the glucose uptake capacity of the whole intestine to change within 24 h?

● No. The properties of the whole intestine are attributable mainly to transporters in epithelial cells near the tips of the villi, which are most exposed to the gut contents. It takes up to 3 days for the new, fully functional cells to reach this point.

Although some cells with the appropriate capacities form within 12 h (Figure 2.20a), they have little opportunity to work when they are located near the crypt. The reverse experiment (Figure 2.20b) pointed to similar conclusions: it took the mice 3 days to acquire the ability to absorb a newly imposed high carbohydrate diet efficiently.

❯ What would happen to the nutrients in the meantime?

● They would be eliminated as faeces, probably with a fair amount of water and the remains of the enzymes that had been produced to digest them.

Not only is the food itself wasted, but digestive enzymes, water and other materials from the mouse's body are lost.

○ Are the data in Figure 2.20 compatible with those shown in Figure 2.15?

● Yes. The time courses of adaptation to dietary change are very similar in both experiments.

○ In the light of these experiments, how would you expect wild mice to respond to novel foods?

● They would eat only a little of a new food, gradually increasing the amounts as the mechanisms of digestion and absorption adjust to process it efficiently.

Experiments show that many wild animals do indeed eat only a little of newly encountered foods, even if to limit themselves in this way entails depleting their storage tissues. Domesticated animals that have been selectively bred to eat readily (and hence to grow as fast as possible) sometimes lack such temperance, and can be induced to eat large quantities of foods that they cannot digest properly, thereby making themselves ill.

### 2.6.1 REDUCING THE COST OF EATING

The experiments described in Section 2.5.1 show that switching from one diet to another entails a significant metabolic cost, but it is difficult to measure exactly how much when so many other biochemical processes are going on at the same time. A more extreme case of similar mechanisms is that of animals that eat large, infrequent meals, or whose food is available only at certain times of the year, and reduce the overall cost of digestion and absorption by partially dismantling the gut between feeds.

The energetic cost of refurbishing the gut to digest food after a long period of fasting has been most thoroughly investigated in snakes, most of which eat other vertebrates as soon as they are large enough to do so. Such prey constitute very concentrated nourishment of almost exactly the same chemical composition as the snake itself, so a few, very large meals thoroughly digested provide adequate nutrition and generate minimal waste. Many of the meals eaten by large snakes amount to 50% of the body mass, and there are records of adult anacondas and pythons swallowing prey that are substantially heavier than themselves.

Even under ideal circumstances, large pythons, boas and vipers eat only about every 1–2 months, and can fast for up to 18 months without coming to any harm. The time courses of the metabolic processes involved in digestion have been studied in the Burmese python (*Python molurus*), a non-venomous species that kills its prey by constriction, feeding mainly on mammals and large birds (Figure 2.21a). The constrictors (boas, anacondas and pythons) are the largest living snakes, and *P. molurus* can grow to a body mass of over 100 kg, but these experimental subjects were young, partially grown specimens weighing less than 1 kg.

Within 2 days of eating a meal of laboratory rats or mice amounting to 25% of their body mass, the snakes' oxygen consumption increased up to 17-fold, and remained continuously high for several days, before gradually declining over the following 2 weeks (Figure 2.21b). This increase in oxygen uptake is greater than that observed when a person or a horse goes from lying down to fast running, and means that glucose and/or fat is being consumed at a very high rate.

The protein synthesis that makes the digestive enzymes, the breakdown reactions themselves and the active transport of the products of digestion all generate additional heat. When fully active, the gut is the largest source of body heat after ATP synthesis in mitochondria.

❯ How would this heat aid digestion?

● Most biochemical reactions, including enzymatic hydrolysis of proteins, fats and carbohydrates, proceed faster at higher temperatures.

So being warmer, whether from the heat produced by digestion or by other means, makes food breakdown faster and often more efficient.

Many poikilothermic (cold-blooded) animals, such as frogs, tortoises and lizards, do not eat unless they are warm enough for digestion to be efficient. But the extra heat produced in the guts can lead to overheating, which can cause permanent damage to delicate organs such as the brain. So large animals, especially those that often have to run fast and/or are exposed to bright sunshine while digestion is in progress, have ways of preventing the excess heat from reaching the brain.

The experimental pythons remained inactive during the period of observation, so the energy must have been used for the synthesis of tissues and secretions needed for digestion and absorption. Examination of specimens killed at various times after feeding showed that much of the extra energy was being devoted to fuelling reconstruction and modification of the internal organs. The stomach, intestine, liver and heart enlarged greatly, starting within a few hours of the prey being eaten, and the gut enormously increased its capacity to digest and absorb nutrients. Figure 2.21c and d show measurements of two aspects of the metabolic response to a meal. Uptake capacity was calculated as the growth of the intestine multiplied by its rate of absorption per unit area of gut, for glucose (Figure 2.21c) or lysine and proline, both products of protein digestion (Figure 2.21d).

❯ From the data in Figure 2.21b–d, could the nutrients obtained from the test meal provide the fuel consumed during the period of high oxygen consumption?

● No. The highest oxygen consumption and the fastest changes in tissue mass and properties occurred before digestion was sufficiently far advanced for significant quantities of any nutrients to be absorbed.

The physiological response to feeding must have been fuelled almost entirely from body stores. The animals need sufficient reserves to rebuild their intestine and other viscera that enable them to digest and absorb the food effectively.

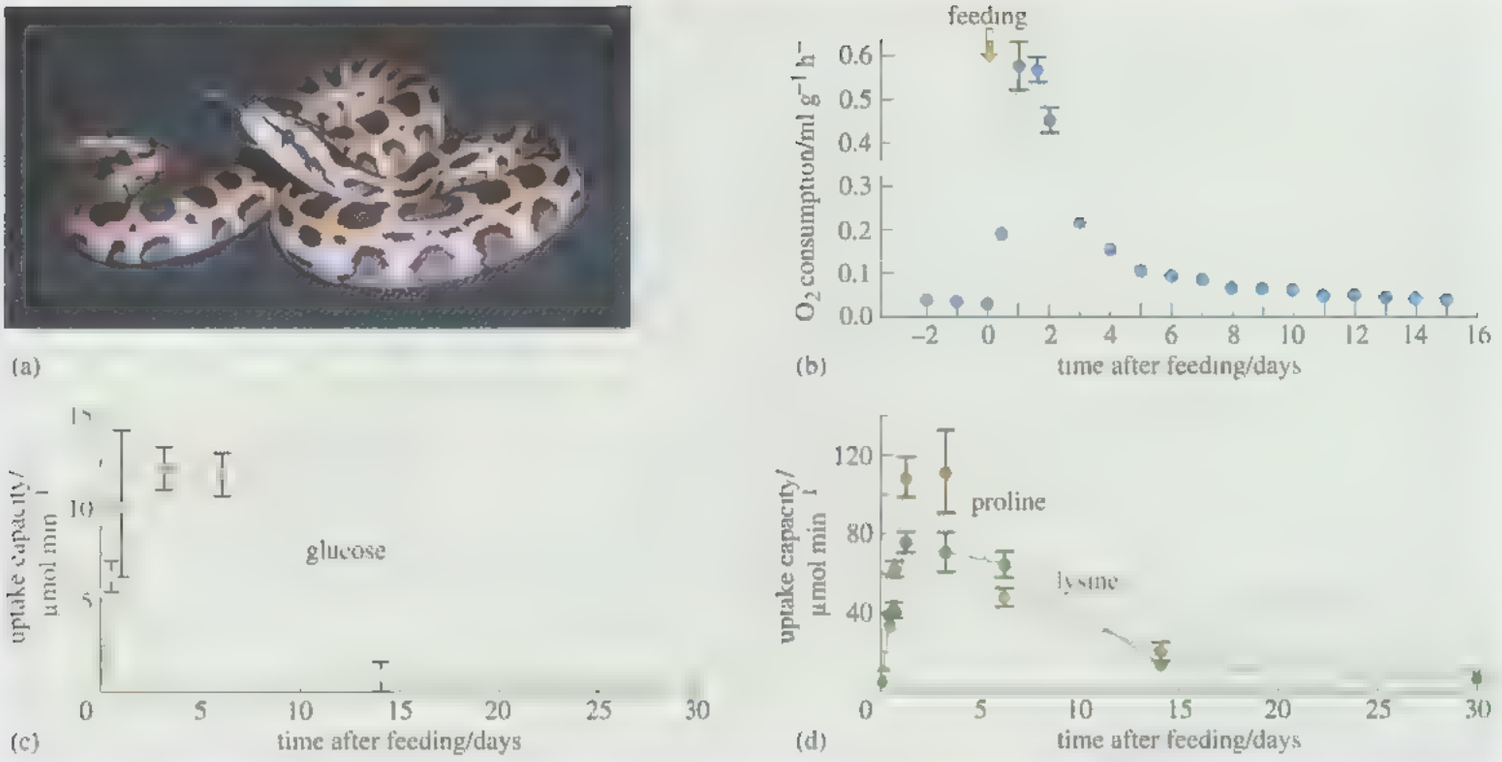
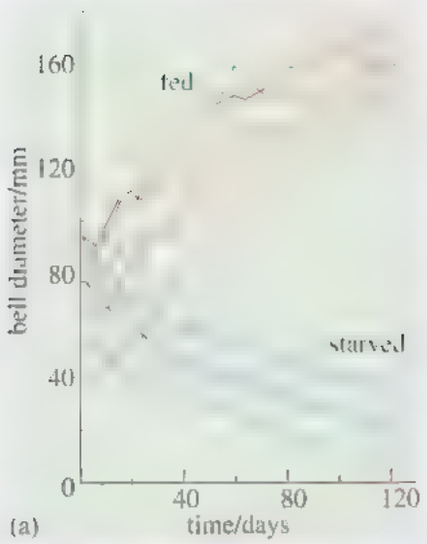



**Figure 2.21** (a) A Burmese python (*Python molurus*). (b)–(d) Physiological changes measured in young Burmese pythons just before and for several weeks after a large meal. (b) Oxygen consumption of the living snakes. Uptake capacities of the whole small intestine for (c) glucose and (d) the amino acid lysine, and proline (another product of protein digestion). Data from Secor and Diamond (1995).

Big snakes apparently alternate brief periods of high energy expenditure during feeding, digestion and absorption, with long intervals in which the gut and other viscera are reduced in size and metabolic capacity. During fasting, the body is maintained at a much lower rate of energy expenditure by body stores. Of course, some energy is expended in converting the nutrients absorbed from the food into storage materials, and in breaking them down again when they are needed, but the quantity is small compared with the cost of maintaining the gut in a state of continuous readiness for frequent, small meals.

Such very long fasts are easier for large snakes than for most other animals because, being cold-blooded, their resting energy expenditure can be quite low. They also travel around very little, and rarely have to move fast, as few predators are likely to tackle them. Although large snakes are often spectacular and generate much fear among potential prey and bystanders (such as ourselves), their efficient digestion and metabolic economies between meals mean that the amount that they actually eat is tiny compared to that needed to sustain birds or mammals of similar size.

During severe illness or other circumstances in which eating is impractical, many other animals close down the gut and rely upon reserves in much the same way as large snakes do. Elderly animals and those weakened by chronic illness or defects of the mouth or gut (e.g. tooth wear, breakage or decay) reach a point beyond which they cannot muster enough energy to fuel digestion. Without the means of obtaining further sustenance, their decline continues until they die, usually, in the case of small animals, within a few hours.




(a)

(b)

**Figure 2.22** (a) Growth during regular feeding in the laboratory and 'degrowth' during starvation in the common jellyfish, *Aurelia aurita*. Each line is a record from a single jellyfish, which was either fed regularly (red) or starved (blue). Data from Hanmer and Jenssen (1974). (b) *A. aurita*. The tissues are almost transparent, except for the four ring-shaped gonads, which are pale pink.

In large animals with substantial energy reserves, the interval between the loss of capacity to feed and death may last some time, long enough to be noticed by human observers. Wild animals that do not die from predation — the fate of the majority — often appear to die of starvation, a conclusion based on the fact that *post-mortem* examination usually reveals an empty gut and little or no adipose tissue. In many such cases, starvation may have been due more to loss of the ability to handle, digest and absorb food than because none was available. The animal 'lost its appetite', but lived on its reserves for days or weeks, or in the case of large snakes, terrapins and tortoises, as long as several months.

## 2.6.2 STORAGE

Fluctuations in the availability of food, and in the animals' capacity to digest and absorb nutrients make some form of storage essential for most free-living animals. Many, in temperate regions most, invertebrates and lower vertebrates feed for only a fraction of the year, often as caterpillars and other kinds of larvae, then survive for long periods eating little or nothing. Although the basic mechanisms of digestion and absorption are fairly uniform throughout the animal kingdom, tissues and processes involved in storing and reclaiming nutrients are quite varied.

Many soft-bodied invertebrates seem to be able to break down almost all their tissues to provide fuel, which supports the metabolism of the surviving cells. Figure 2.22a shows some data from *Aurelia aurita*, a small jellyfish that occurs in surface waters worldwide and is often found stranded at low tide on beaches around Britain (Figure 2.22b).

❯ Which of the two processes measured in Figure 2.22a is the faster?

● Growth: the jellyfish grow faster when fed frequently than they shrink in starvation.

The growth rate observed in these specimens may be close to the maximum, and might be much slower if less food were available. Even medium sized individuals can sustain themselves for at least four months by reclaiming their tissues as fuel. These jellyfish can shrink to less than 10% of their former size and still remain active, because their simple body structure works well over a wide range of sizes. Many other soft-bodied invertebrates respond similarly to starvation, but this strategy would not be appropriate for arthropods and vertebrates. Their hard skeletons of cuticle or bone are mostly unreclaimable, so overall body dimensions cannot shrink very much. Severely wasted muscles become too weak to support the skeleton or to move the limbs or jaws efficiently, and, as explained in Section 2.6.1, depleted guts cannot deal with food properly. These more complex animals have specialized storage tissues which can sustain the animal for some time before depletion of the general body fabric (as in Figure 2.22a) becomes necessary.

The main storage tissue in larger and more complex arthropods is the **fat-body**. Its physiology has been most thoroughly studied in insects, where it stores glycogen, lipids, and, at least in juveniles, proteins. It also performs many of the

metabolic functions carried out by the vertebrate liver. Glycogen and lipids are insoluble in water, so can accumulate in large quantities inside cells without promoting the uptake of too much water. The fat-body can become huge (up to 65% of the body mass) in caterpillars and grubs just before they pupate and are transformed into adult insects. In locusts or migratory butterflies about to set off on long flights, the fat-body can be the most conspicuous organ in the body, enveloping the gut, flight muscles and excretory organs.

Most kinds of vertebrate cells (neurons are the best-known exception) store small quantities of glycogen and/or small droplets of triacylglycerols as an energy reserve for their own use. Metabolically active tissue such as muscle and liver can release small amounts of protein to supply materials for rebuilding damaged tissues and for the formation of eggs, but tissue proteins are usually broken down to form glucose only in advanced starvation. Liver and muscle take up excess glucose from the blood and store it as glycogen, which is converted back to glucose and used by these and other tissues during brief periods of fasting. However, as Figure 2.23 makes clear, the quantities are quite small, usually no more than 5–7% of the tissue mass.

❯ Why can't these tissues hold more storage materials?

● Because their cytoplasm is devoted to other functions, e.g. movement in the case of muscle. The necessary enzymes and structural proteins would be disrupted by large quantities of storage materials.

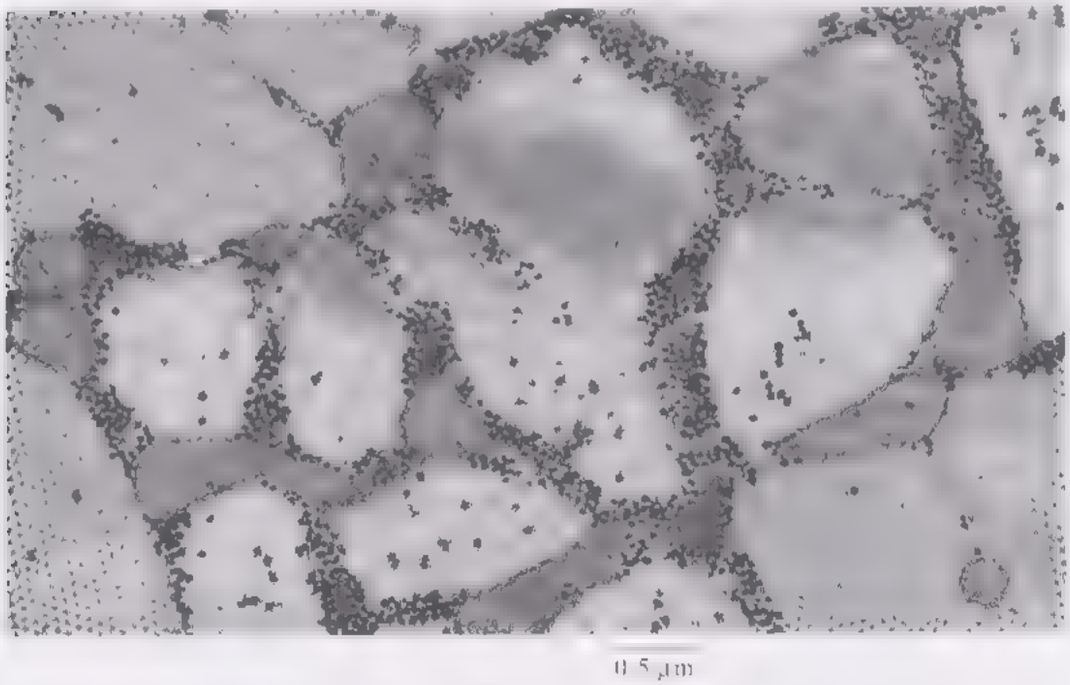



**Figure 2.23** Electron micrograph of a fragment of the gastrocnemius (calf) muscle of a person who has just eaten a carbohydrate-rich meal. The muscle took up the glucose and converted it into insoluble glycogen, which forms many small granules, here stained dense black. This picture also shows several mitochondria, which appear as round structures, filled with membranes. The mitochondria and the glycogen granules are packed around the contractile components of the muscle, which are seen as arrays of grey spots in this cross-section.

**Figure 2.24** Adipose tissue from a guinea-pig. Note the round outline of the adipocytes visible in this unstained whole-mount. The yellow colour is that of the storage lipids.

Some fish, especially sharks and rays (class Chondrichthyes), may accumulate large quantities of fat in the liver, but all terrestrial vertebrates sequester most of their storage lipid in adipose tissue. This tissue consists almost entirely of adipocytes (Figure 2.24), cells that are specialized for storing lipid to be used by other body tissues. Being insoluble, triacylglycerols can be stored in an almost pure form in the single large droplet that fills most of the adipocyte.

❯ What would be the advantages of having a specialized storage tissue?

● Lipids can be safely stored in far larger quantities than if they were 'loose' in the blood or in tissues such as muscle or liver which are specialized to other functions.

The quantities of adipose tissue that can be safely maintained are limited more by animals' capacity for carrying the tissue around than by metabolic factors.

Adipocytes expand by taking up long chain fatty acids from the chylomicrons, or, especially in herbivores, by synthesizing them from glucose and other small molecules. They are unusual among animal cells in being able to undergo enormous changes in volume, up to 100-fold in some species. The adipose tissue of birds just before migration or of mammals about to enter hibernation may be up to two-thirds by mass triacylglycerols. Naturally obese animals may also have proportionately more adipocytes, perhaps two or three times as many as in continuously lean species of the same size. Polar bears are top carnivores, feeding almost entirely on seals in the Arctic, where the adults have no natural predators. They readily fatten when food is available and can fast for months while seeking prey, often travelling long distances between hunting grounds. Pregnant females may exceed 50% by mass adipose tissue before they retire inland to a maternity den where they give birth and suckle twins (sometimes triplets) for three months or more without eating anything themselves.

The arrangement of adipose tissue in most fish, amphibians and reptiles is similar to that of arthropod fat-bodies, i.e. a few large masses, usually in the abdomen, but that of mammals is partitioned into many depots spread around the body. Recent research reveals that adipocytes have site specific properties and some are equipped to respond to local signals from the immune system, as well as, or instead of, the hormones secreted during fasting. Some (but probably not all) adipocytes also contribute to the regulation of food intake by secreting signal molecules into the blood that act directly on parts of the brain that control appetite. Adipocytes should be regarded as 'managers' of metabolic stores, rather than simply as a repository for them: they synthesize special enzymes, carrier proteins and receptors that enable them to take up, hold and release their stores in response to hormones and other signals released from other cells.

Small quantities of many vitamins and minerals are also stored in various body tissues, but amounts are very limited, often only sufficient to last for a few weeks, because almost all can be harmful in excess. As described in Section 2.5.2, iron is toxic at high concentrations, so, in spite of its physiological importance to vertebrates, it can only be stored bound to special proteins in the liver.

## SUMMARY OF SECTION 2.6

1 The lining of the mammalian intestine is replaced every few days as new epithelial cells mature and pass along the villi, where they are sloughed off. The absorptive capacities of newly formed cells are adapted to changes in diet.

2 Enzyme secretion and active transport are energetically expensive. Animals that eat large nutritious meals very infrequently may save energy by partially dismantling the gut while fasting. Reassembling it after feeding requires high rates of energy usage sustained for many hours.

3 Rebuilding the gut and digestion release large quantities of heat, which speeds up digestion and metabolism but may overheat other tissues when in excess.

4 Many animals can reabsorb and break down their tissues during starvation but most vertebrates and more complex invertebrates have specialized storage tissues.

5 Glycogen is a short-term energy store often used by the tissue that sequesters it.

6 The arthropod fat-body and vertebrate adipose tissue are specialized for storing lipid for use by other tissues. These tissues can become massive, enabling animals to fast for long periods and/or travel long distances. Mammalian adipose tissue is split into many depots, some of which have specialized site-specific properties.

7 Most other nutrients are too toxic to be stored in large quantities.

## CONCLUSIONS

Natural selection usually acts to thwart rather than promote the transfer of nutrients from plant to animal or from prey to predator, so there is no permanent, ideal solution to the problem of obtaining food. Obtaining and digesting food are central to the structure and habits of all heterotrophs. Continuous adaptation of diet, feeding habits, digestion and storage lead to diversification of all lineages of organisms. Digestive mechanisms determine not only their diet but also other aspects of their biology, such as the amount of time they spend feeding, their capacity for exercise, or their metabolism, including energy metabolism and (in mammals) the composition of the milk they secrete.

## REFERENCES

Bertsch, A. (1984) Foraging in male bumblebees (*Bombus lucorum* L.): maximizing energy or minimizing water load?, *Oecologia*, **62**, pp. 325–336.

Diamond, J. M. and Buddington, R. K. (1987) Intestinal nutrient absorption in herbivores and carnivores, in *Comparative Physiology: Life in Water and on Land*, P. Dejours, L. Bolis, C. R. Taylor, and E. R. Weibel, (eds), pp. 193–203.

Ferraris, R. P. and Diamond, J. M. (1993) Crypt/villus site of substrate-dependent regulation of mouse intestinal glucose transporters, *Proceedings of the National Academy of Sciences, USA*, **90**, pp. 5868–5872.

Hanmer, W. M. and Jenssen, R. M. (1974) Growth, degrowth, and irreversible cell differentiation in *Aurelia aurita*, *American Zoologist*, **14**, pp. 833–849.

Secor, S. M. and Diamond, J. (1995) Adaptive responses to feeding in Burmese pythons: pay before pumping, *Journal of Experimental Biology*, **198**, pp. 1313–1325.

Slater, A. F. and Cerami, A. (1992) Inhibition by chloroquine of a novel haem polymerase enzyme activity in malaria trophozoites, *Nature*, **355**, pp. 167–169.

Vollrath, F., Downes, M. and Krackow, S. (1997) Design variability in web geometry of an orb-weaving spider, *Physiology and Behavior*, **62**, pp. 735–743.

## FURTHER READING

Pond, C. M. (2000) *The Fats of Life* (2nd edn) Cambridge University Press. [This book explains in simple language the biology of feasting and fasting, fattening and slimming in wild animals as well as people.]

Pough, F. H., Janis, C. M. and Heiser J. B. (1996) *Vertebrate Life* (5th. edn), Prentice Hall, New Jersey. [A wide-ranging review of the diversity, function and evolution of vertebrates. (6th edn expected 2002)]

Schmidt-Nielsen, K. (1997) *Animal Physiology* (5th edn) Cambridge University Press. [A useful textbook of comparative physiology, with emphasis on vertebrates, packed with good ideas and useful, accurate facts, but rather poorly written.]

Stevens, C. E. and Hume, I. D. (1995) *Comparative Physiology of the Vertebrate Digestive System*, Cambridge University Press. [A beautifully illustrated account of physiological and ecological aspects of diet and digestion.]

# GENETIC DIVERSITY

## 3.1 INTRODUCTION

In Chapter 1 of this book we discussed some of the ways in which plants and animals are adapted to the fluctuating nature of their environment, and in Chapter 2 considered diverse ways of dealing with food. How does such diversity arise? This chapter addresses the question by firstly examining how natural selection acts on differences in the phenotypes of organisms. It then moves on to consider the basis of variation in sexually reproducing organisms and how heritable variation arises. You will be able to draw on this information when comparing the costs and benefits of sexual and asexual reproduction in Chapter 4, and the long-term evolutionary consequences of asexual reproduction. Throughout Chapter 3 you will be presented with opportunities to consider how populations of species may change and some of the circumstances in which these changes result in evolution.

## 3.2 NATURAL SELECTION

Clearly, not all organisms do survive and go on to reproduce, so why should some succeed whilst others fail? We can begin to consider the reasons by first examining the four basic premises of the theory of evolution by natural selection. These principles were outlined by Charles Darwin in his introduction to *The Origin of Species by Means of Natural Selection*, following his joint communications with Alfred Russel Wallace in 1858:

1 More individuals are produced than can survive.

2 There is a *struggle for existence* because of the disparity between the number of individuals produced in reproduction and the number that can survive.

3 Individuals show *variation*. No two individuals are exactly the same. Those variants with advantageous features have a greater chance of survival in the struggle (*natural selection*).

4 As these selected variants produce offspring similar to themselves (*the principle of inheritance*), they become proportionally more abundant in subsequent generations.

**Natural selection** favours those variants in a population that leave the most offspring, who, by virtue of inheriting the same characters from their parents, should likewise be favoured by natural selection. Natural selection selectively multiplies variants between one generation and the next, so we can evaluate its effect on any particular variant by measuring how many offspring it leaves. Of course, for the offspring to reproduce themselves, they must survive to sexual maturity, so here are the two components for the recipe for success: survival (also known as viability) and reproduction (also known as fecundity).

Consider two kinds of rabbits: those with black coats that produce 20 young of which, on average, 50% survive to adulthood, and those that are grey and produce 18 young, of which 60% survive.

› Assuming that these differences are inherited, what kind of rabbit is favoured by natural selection? To answer this question, multiply the number of young produced by each rabbit type by its survival to adulthood.

● On average, black rabbits produce $0.5 \times 20 = 10.0$ successful offspring surviving to adulthood, and grey rabbits produce $0.6 \times 18 = 10.8$ successful offspring surviving to adulthood. Therefore the grey rabbits are favoured over the black.

The product of survival and reproduction is the measure of the **fitness** of a variant, in this case the black or grey phenotype of rabbit. The outcome of natural selection is determined by the relative number of descendants left by different phenotypes, so fitness is conventionally measured with respect to the superior phenotype. The phenotype with the highest relative fitness is assigned a fitness value of 1 and the relative fitness of other phenotypes is given as a frequency value relative to this.

If the fitness of grey rabbits is 1, what is the relative fitness of the black ones?

● The relative fitness of the black rabbits is calculated relative to the fitness of greys (which are assigned the value 1 as they have a higher fitness than the black rabbits).

reproductive success of blacks = 10.0

reproductive success of greys = 10.8

so relative fitness of blacks $= \dfrac{10.0}{10.8} = 0.93$

In percentage terms, this calculation means that the fitness of black rabbits is 93% of that of greys. Does a difference as small as 7% matter? Such a difference may seem slight, but even small differences in fitness sustained over several generations can cause a significant evolutionary change in the relative proportions of the phenotypes in a population.

› How many descendants could one black rabbit have over four generations? How does this number compare with the number of descendants from a grey rabbit over the same number of generations?

● After four generations, a black rabbit has $10 \times 10 \times 10 = 10\,000$ descendants.

A grey rabbit has $10.8 \times 10.8 \times 10.8 \times 10.8 = 13\,603$ descendants

$\left(\dfrac{3603}{10\,000} \times 100 = 36\%\text{ more than the black}\right)$.

So for a rabbit population that started off with a 50 : 50 ratio of grey : black types, a 7% difference in relative fitness can result in a ratio shift to 58 : 42 in as little as four generations. Typically, a rabbit population explosion would exhaust the food supply and cause an increase in mortality and a decrease in fecundity. These effects are what Darwin meant by the struggle for existence caused by the disparity between the number of offspring that can be produced and the number that can actually survive.

Recall from Chapter 2 that animals have many ways of improving the efficiency of dealing with food, including adaptations for catching and digesting prey, developed in response to selection pressures to reduce the costs of eating.

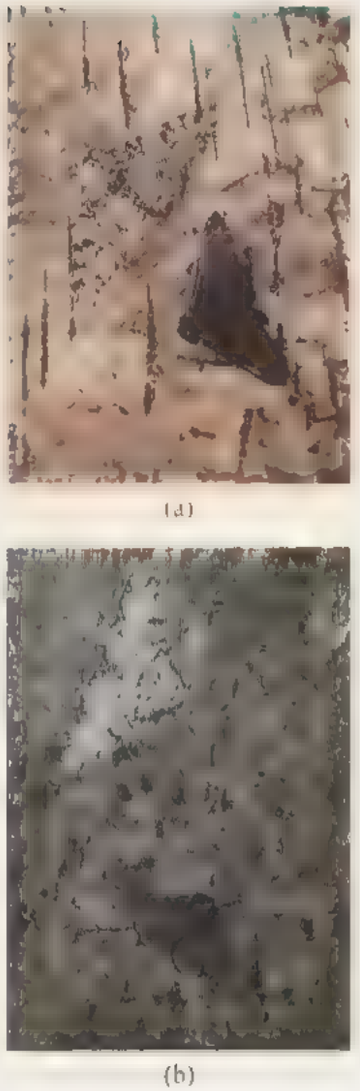
Note that if an organism produces a large number of offspring that survive to reproductive age, but fail to reproduce (perhaps because they are sterile), then the genetic make-up of the ancestral organism is not passed on to any future generations.

## 3.3 The genetic basis of variation within species

Within a population of a species there can be great genetic variation which may be immediately apparent in morphological characters. For example, in the peppered moth, *Biston betularia*, (Figure 3.1) dark coloured, *Biston betularia carbonaria*, moths have at least one copy of the *T* allele (the dominant allele, conferring blackness) whilst pale coloured, *Biston betularia typica*, moths have two copies of the recessive, pale colour allele, *t*.

Alternative characters in garden peas, such as pod shape (inflated or constricted), pod colour (green or yellow), seed shape (round or wrinkled), and flower colour (purple or white), are directly related to single pairs of alleles.

The shell of the land snail, *Cepaea nemoralis*, may be brown, pink or yellow. The colours are determined by a locus with six alternative alleles. An individual snail has two alleles at a locus, but with more than two alternatives in the population, there are more possible combinations than for the garden pea examples above, with only two alternative alleles described. The shell colour loci are closely linked to others at which the alleles control the presence and appearance of up to five bands running round each whorl of the shell. In addition, there are six unlinked genes, each with two alleles per locus, which determine other aspects of band production including intensity, colour and suppression (absence) of some bands, and colour of the body. The morphology of an individual snail shell is determined by the interactions of alleles at many loci and Figure 3.2 overleaf shows just three of the possible phenotypes.



(a)

(b)

**Figure 3.1** The peppered moth, *Biston betularia*. The typical form and the *carbonaria* form seen against the bark of a tree from (a) a rural area, and (b) an industrialized area. The relative abundances of the two forms have altered with changes in the levels of atmospheric pollution.

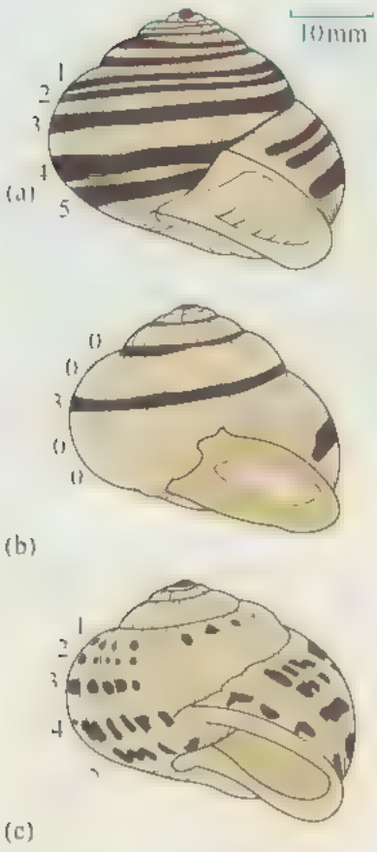



**Figure 3.2** Examples of shell patterns in *Cepaea nemoralis*: (a) five-banded shell; (b) shell with bands 1, 2, 4 and 5 suppressed; (c) shell with five punctate bands.

**BOX 3.1 USE OF GENETIC TERMS**

Use of the terms 'locus', 'allele', 'gene' and 'character' can sometimes be confusing. The location of a gene on a chromosome is termed a 'locus' (plural, loci). Each gene can exist in one or more forms and strictly we should use the term 'allele' as a form of a gene. However you are likely to find 'allele' and 'gene' used interchangeably in some situations as in 'allele for blackness' and 'gene for blackness' in the peppered moth. The forms (alleles) of a gene are genes themselves. Notice also the association between 'gene' and 'character' (or characteristic) as in gene for the character 'seed shape' in peas. From the *Cepaea* example above, you can see that alleles at more than one locus, and hence more than one gene, may be involved in the expression of a character such as shell banding, resulting in a mid-banded shell as in Figure 3.2b.

During meiosis, the genome is restructured through independent assortment of chromosomes and crossing over between homologous chromosomes, resulting in a huge number of possible allele combinations in the gametes (discussed further in Section 3.4). The possibilities for different allele combinations at fertilization are also vast, resulting in a huge range of possible phenotypes such as shell colour and banding patterns in *Cepaea*.

Much genetic variation gives rise to differences in enzymes that cannot be recognized from morphological differences, so this genetic variation must be investigated using molecular techniques such as gel electrophoresis, immunology, and sequencing of proteins and genes.

### 3.3.1 THE GENE POOL

The numbers of different alleles of genes form the **gene pool** for a population of organisms. Changes in the gene pool take place as a result of natural selection and also mutation and genetic drift. Genetic drift will be considered later in this section. At the genetic level, evolution is expressed as change over time in the frequencies of alleles in populations. Remember that these alleles are expressed through the phenotype of the individuals carrying them and that natural selection acts on phenotypes.

Differences in fitness among phenotypes lead to changes in the gene pool of a population, as the less fit phenotypes and their alleles are reduced in frequency by natural selection. Natural selection also acts, via the phenotype, on new alleles produced by mutations and new combinations of alleles arising from recombination during meiosis.

It is generally easiest to study genetic systems that involve one locus and a pair of alleles. When there are two or more alleles at a locus, the character is said to be **polymorphic** (as in the shell colour for *Cepaea nemoralis*, above). However, the inheritance of the majority of characters depends on several loci, often with more than one allele at each. Then the character is said to be **polygenic** (as in the overall appearance of the *Cepaea* shell). Such polygenic characters may have many different forms (recall that in *Cepaea* there are many different shell colour and banding patterns that are determined by the interactions of several genes) or

they may vary continuously so that the ratio of different phenotypes in a population is best represented by a distribution curve. Shifts in the frequencies of phenotypes may reflect a shift in the frequencies of many alleles.

Selection acting against extreme variants and in favour of the most common phenotype is termed **stabilizing selection** as it tends to maintain the underlying genotype frequencies within their existing distribution. Such selection does not bring about change and hence does not lead to evolution. **Directional selection** causes the phenotype distribution to change towards one extreme as a result of selection against the other extreme, whereas **disruptive selection** leads to a change in phenotype distribution towards both extremes as a result of selection against the intermediate phenotype (Figure 3.3).

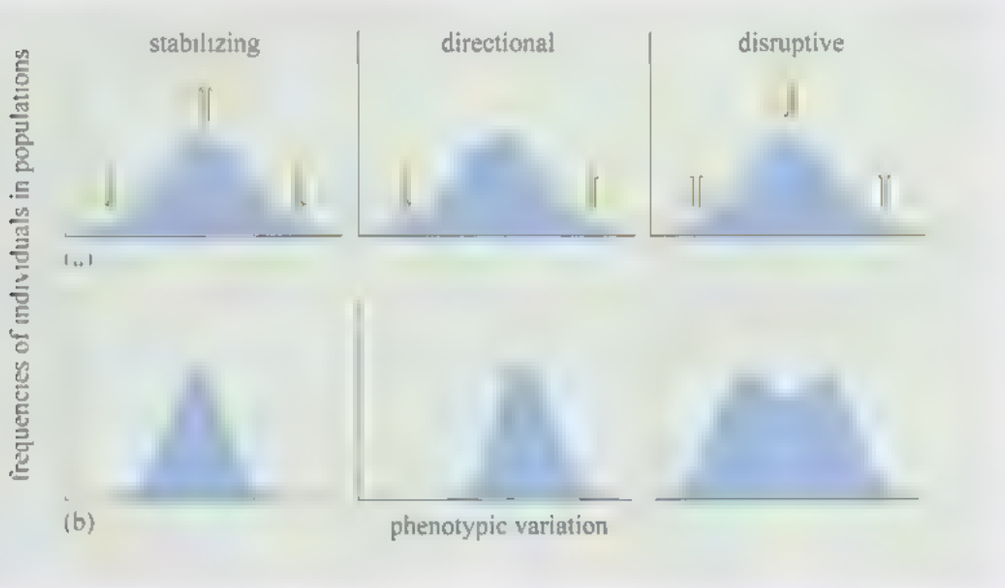



**Figure 3.3** The effect of stabilizing, directional and disruptive selection on the distribution of phenotypes in a population: (a) shows the distribution as selection begins, and (b) shows the distribution after selection.

Stabilizing selection has been investigated in South American butterflies of the genus *Heliconius*. *Heliconius erato* and *Heliconius melpomene* are both distasteful to birds and deter attempts at predation with warning colourations. The colour patterns of their wings may differ at different localities, but at any one locality the wing patterns of the two species are more similar to one another (Figure 3.4). It has been proposed that this situation may represent an extended system of **Müllerian mimicry** — the mimicry (resemblance strategy) which afford unpalatable species increased protection from predators. (This topic is discussed further in Chapter 5.)

- What would this hypothesis suggest about the fate of poor mimics whose wings are dissimilar from those of distasteful butterflies?

- Butterflies that deviate from the usual pattern, and are therefore poor mimics, are more likely to be caught and eaten than other individuals. There is natural selection against them.

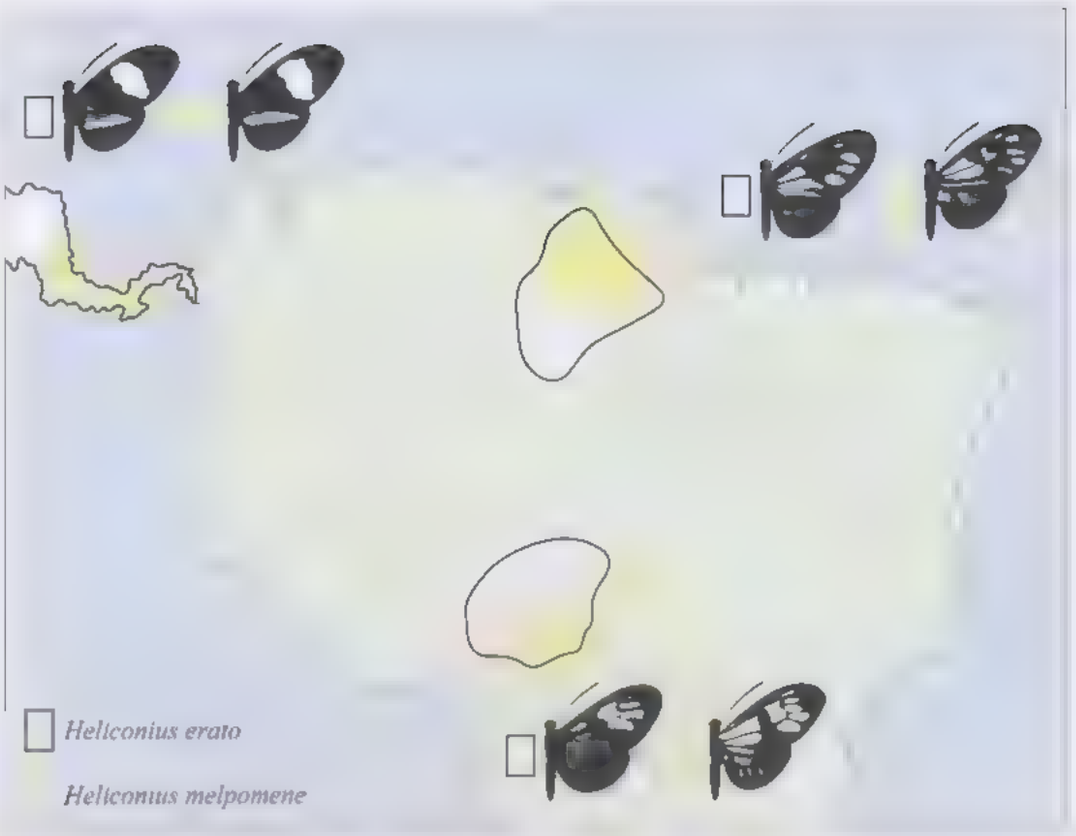



**Figure 3.4** Parallel variation in three races (populations within one species that are phenotypically distinct) of *Heliconius melpomene* (yellow areas) and *Heliconius erato* (white areas). Very similar wing patterns for the two species are found in the same general areas.

Experiments were conducted in which the wing patterns of *Heliconius erato* were altered by applying paint to the wings (with controls having the same amount of paint applied to their wings without altering the pattern). Under stabilizing selection, poor mimics would be more likely to be caught and eaten by birds, and experiments confirm this prediction. In addition, when races of a species with different wing patterns meet, there are very narrow zones occupied by both parental species. It has been suggested that stabilizing selection for the parental patterns leads to the elimination of the first generation of hybrids. These hybrids would be less fit as they are killed more often by birds.

A case of sympatric speciation, where the two parent species occupy the same geographic area, has been described, based on a model of disruptive selection. Two species of lacewings, *Chrysopa carnea* and *Chrysopa downesi*, insects of the order Neuroptera, occur in the same geographic area in the northeastern United States. *Chrysopa carnea* inhabits grasslands and meadows and produces several generations between late spring and summer every year. During the reproductive period, the adults are pale green in colour and match the light green foliage. At the end of the summer, the adults enter a resting stage, turn from pale green to reddish brown and move to the senescent foliage of deciduous trees. *Chrysopa downesi* lives in coniferous woodland and breeds once a year in early spring. The adults are dark green in colour throughout the year and are camouflaged amongst the conifers. Therefore in their natural environment, the two species are reproductively isolated by differences in habitat and seasons of reproduction. However, under laboratory conditions they can hybridize freely and produce fully viable offspring.

In hybridization experiments, it was found that *Chrysopa carnea* was homozygous for the allele designated *G1*, giving a light green phenotype whereas *Chrysopa downesi* was homozygous for the allele, *G2*, which produces the dark green colour. Heterozygotes, *G1 G2*, were intermediate in colour.

This investigation suggests that disruptive selection favoured homozygotes in different habitats — *G1 G1* in meadows and *G2 G2* in coniferous woodland. The heterozygote, *G1 G2*, is selected against because of its reduced protective colouration in both habitats. Habitat separation led to assortative mating (like phenotypes mating with like). Subsequently there was selection for differences in reproductive timing. Reproductive isolation was therefore complete, resulting in two species where there had previously been one.

Directional selection has been seen in populations of the peppered moth, *Biston betularia*, a moth that is active at night and settles during the day on tree bark which is usually covered with pale coloured lichens (see Figure 3.1). In the early part of the 18th century in Britain almost all moths were the pale *typica* form, with the genotype of almost all individuals being *t t*. When a mutation produced a *T* allele (the dominant allele conferring blackness), these dark moths were very strongly selected against by bird predation as they were highly visible on lichen-covered tree bark. However, as the air in some environments became dirtier during the Industrial Revolution, the lichen was killed by atmospheric pollutants. Dark coloured *carbonaria* moths were better camouflaged on tree bark no longer covered by lichen, and so the frequency of *carbonaria* phenotypes increased. There was directional selection from *typica* towards *carbonaria* phenotypes.

- If an industrial environment subsequently becomes much cleaner, what do you predict to be the action of selection on the phenotypes of *Biston betularia*?

- Lichen growth would occur again on the trunks and the direction of selection would reverse, with the frequency of *typica* phenotypes increasing and that of *carbonaria* decreasing.

This change in the direction of selection has indeed been recorded and demonstrates that selection can proceed in different directions at different times in the same location. The direction of selection may also differ in different places at the same time if the environments vary.

### 3.3.2 GENETIC DRIFT

In real populations, allele frequencies tend to fluctuate from generation to generation. If the population is large, these fluctuations are relatively small, but fluctuations in small populations may have relatively much larger effects on allele frequencies. These random fluctuations in allele frequencies are known as **genetic drift**.

Genetic drift is the consequence of the sampling process that occurs in reproduction. When gametes are formed, each haploid gamete receives, at random, one of the two alleles that the individual had previously inherited from its parents. In addition, only a small fraction of gametes actually succeeds in uniting through fertilization (the process of **syngamy**) and developing into organisms of the next generation. In very small populations, genetic drift can have a strong effect if, by chance, one individual produces more offspring than others, so its alleles constitute a greater proportion in the next generation. A particular allele could disappear from a small population by genetic drift.

- What is the effect of genetic drift on the genetic variation within a small population?

- The genetic variation is likely to be reduced as a result of chance loss of alleles.

Genetic variation between populations may arise also when a small number of individuals from an originally large population become founders of a new population, as could happen as a result of emigration to a new habitat, or as a consequence of a natural catastrophe such as drought or volcanic eruption, which wipes out much of the original population or leaves small populations geographically separated from one another. In such circumstances, the allele frequencies of the new population are those of the founders, which may or may not be representative of the original population. This sampling effect, due to small numbers, is known as the **founder effect**.

If an allele, rare in the original population, happens to be among the founders, what can we predict about its future frequency in the founder population in comparison with the original population? Assuming that it is not eliminated by genetic drift while the population is very small, the rare allele has a much greater chance of becoming common than it had in the original population.

Evidence from human populations shows that alleles present in small founding numbers of individuals may have disproportionately large frequencies of alleles that are rare in the parental population from which they were derived. The Afrikaner population of South Africa is mainly descended from a group of immigrants that landed in 1652, with almost one million living Afrikaners bearing the names of 20 original settlers. Some of these colonists carried a number of rare genes, one being that for Huntington's disease, a lethal autosomal dominant gene whose effects become manifest in middle age. Most cases of Huntington's in the modern Afrikaner population can be traced back to one original colonist.

The Afrikaner population also has a much higher frequency of the dominant allele that causes a condition known as porphyria variegata than their parent population in the Netherlands. Carriers of this allele suffer a severe, sometimes lethal, reaction to barbiturate anaesthetics. In the absence of exposure to barbiturates, the porphyria variegata allele appears to be a neutral mutation (i.e. neither selected for nor against), but the relative fitness differences of carriers and non-carriers becomes evident when the environment changes to contain (for some individuals) barbiturates.

- Could subsequent evolution within small populations be only as a result of chance effects such as genetic drift and mutation?

- No, the random effects of genetic drift and mutation within a founder population are important, but if there are fitness differences between individuals in the population, natural selection also takes place.

A founder population that colonizes an environment previously unoccupied by other members of that species (e.g. an island) contains a fraction of the genetic diversity in the gene pool of the parent species and all subsequent evolution proceeds from this limited genetic diversity. Mutations that are disadvantageous as heterozygotes but favourable as homozygotes have a greater chance of being incorporated in a small population. If these homozygotes are better adapted to the new environment, they are rapidly selected for, and divergence from the parent species occurs. Divergence is likely to take place much faster in a small founder population, so speciation can proceed more rapidly. Such a form of allopatric speciation (where the two populations are geographically separated) is considered to have taken place on the Hawaiian Islands, where, for example, extensive studies of the many species of *Drosophila* suggest that all the endemic species may have descended from a single fertilized female, probably from South America, colonizing the oldest islands.

### SUMMARY OF SECTIONS 3.2 AND 3.3

1 The product of survival to reproductive maturity and reproduction (viability and fecundity) is the measure of fitness of a phenotype.

2 Natural selection acts on the different phenotypes within a population. It can only affect the course of evolution if these differences are heritable, i.e. produced by different genes.

3 Random fluctuations in allele frequencies occur in populations and genetic drift is likely to have larger effects in small populations.

4 The founder effect describes the effect on genetic variability of the small size of a colonizing population, which might consist of one or a few individuals. Such a population can never contain more than a fraction of the total genetic variability of the parent population. An allele that is rare in the parent population has a chance of becoming common if it is present in a founder population, provided that it is not eliminated by genetic drift when the population is small.

## 3.4 SOURCES OF GENETIC VARIATION

Let us now consider some of the sources of the genetic variation that we have been discussing.

In sexually reproducing organisms, **recombination** through crossing over and independent assortment of alleles during meiosis leads to genetic variability between parents and offspring. Let us first consider how variation arises through independent assortment. Study Figure 3.5.

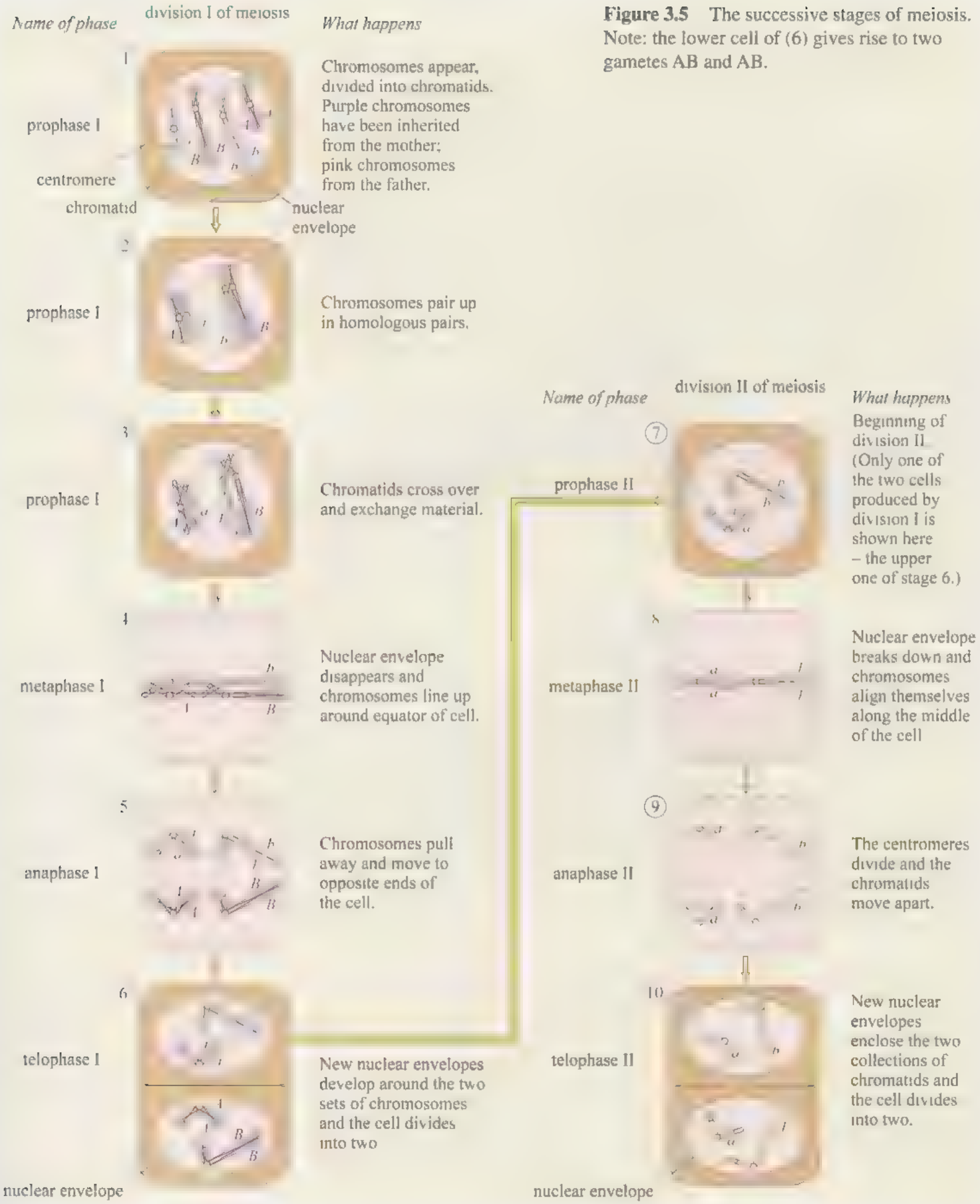



**Figure 3.5** The successive stages of meiosis. Note: the lower cell of (6) gives rise to two gametes AB and AB.

- Metaphase 1 in Figure 3.5 shows one possible arrangement of the two pairs of chromosomes. What other arrangements are possible?

- Three other arrangement are possible. Figure 3.6 shows all four ways in which two pairs of chromosomes could align in a cell at metaphase I, with the arrangement in Figure 3.5 represented in Figure 3.6b (simplified, with no chiasmata shown between homologous pairs).

- What consequences would these alternative arrangements have on the genotypes of the resulting gametes?

- The outcome of meiosis in Figure 3.5 is two types of gametes with respect to the alleles considered. However an alternative orientation of one pair of homologous chromosomes results in two different types of gametes (Figure 3.6c or d). Notice the outcome of the orientation of chromosomes in Figure 3.6a is the same as that in Figure 3.6b. The four arrangements are equally likely, so this independent assortment leads to gametes with four different genotypes, deviations from a 1 : 1 : 1 : 1 ratio being by chance alone.


**Figure 3.6** The genotypes of gametes resulting from each of the four different arrangements of chromosomes at metaphase I of meiosis.

Therefore an organism heterozygous for two pairs of alleles (i.e. alleles at two loci) situated on different chromosome pairs (i.e. non-homologous chromosomes) produces four different types of gamete in approximately equal numbers.

Let us now consider a cross between two maize plants heterozygous for grain colour and grain shape. Purple grain colour is dominant to white and smooth grain shape is dominant to wrinkled.

› What are the phenotypes of grains from which the heterozygous parent plants developed?

● As the plants are heterozygotes, they must have developed from grains with the dominant characters purple and smooth.

Choosing the symbols *G* for purple grain, *g* for white grain, *T* for smooth grain and *t* for wrinkled grain, Figure 3.7 shows a Punnett square, a way in which the genotypes of gametes and the genotypes of the resulting offspring of the cross between plants heterozygous for both pairs of alleles are recorded. It also includes the phenotypes of the resulting maize grains.

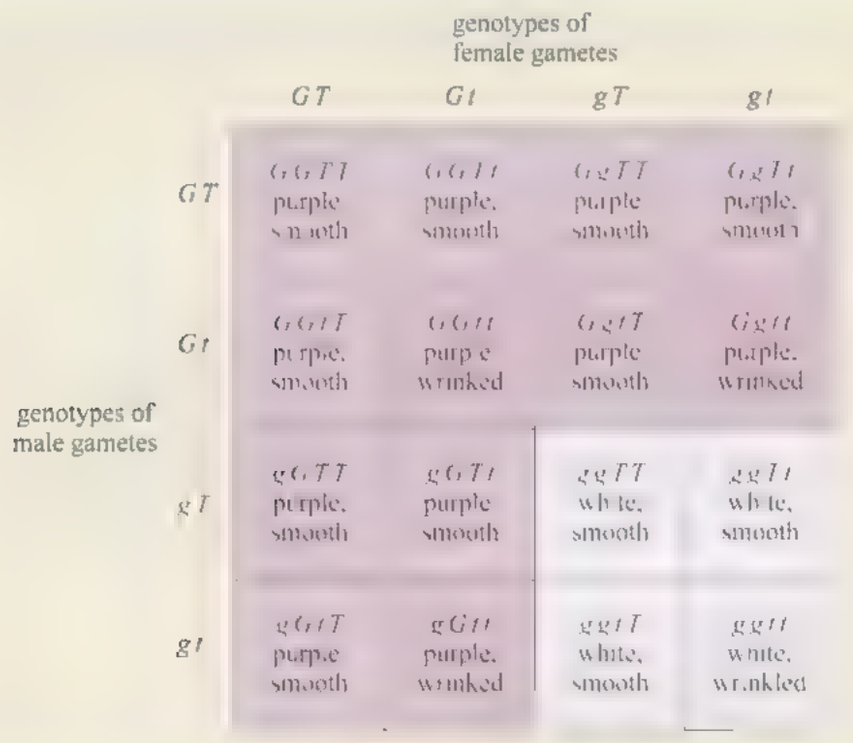



**Figure 3.7** Punnett square showing the result of a cross between two maize plants heterozygous for two pairs of alleles. Alleles: *G*, purple grain, *g*, white grain, *T*, smooth grain and *t*, wrinkled grain.

- Both parent plants developed from purple grains that were smooth. What do you notice about the phenotypes of the offspring grains?

- In addition to purple and smooth, some of the offspring grains are white and wrinkled; others white and smooth; others purple and wrinkled — three of the phenotypes are different from the parent grains.

Independent assortment of alleles during meiosis, followed by fertilization, results in expression of the recessive characters in some of the offspring.

- How many different genotypes are there in the offspring?

- There are nine genetically different types of offspring.

- How many different phenotypes are produced in the offspring?

- There are four different phenotypes.

- In what ratios do these offspring phenotypes occur?

- Counting the number of squares in which a phenotype occurs in Figure 3.7 gives a ratio of 9 purple, smooth : 3 purple, wrinkled : 3 white, smooth : 1 white, wrinkled.

A cross between plants heterozygous for two pairs of alleles carried on different chromosome pairs gives this 9 : 3 : 3 : 1 ratio of phenotypes in the offspring, with expression of phenotypes not seen in the parents. The random nature of independent assortment and the processes that determine which gametes actually achieve fertilization mean that the ratios observed in a breeding experiment may differ from an exact 9 : 3 : 3 : 1 relationship.

When we start to consider more than two pairs of heterozygous alleles the situation rapidly becomes more complex. If both parents are heterozygous for three pairs of alleles (at three loci) on three different chromosome pairs, then each produces $2 \times 2 \times 2 = 8$ genetically different types of gamete and $3 \times 3 \times 3 = 27$ different genotypes amongst their offspring. To generalize, if parents are heterozygous for $N$ different pairs of alleles ($N$ loci), all on different chromosome pairs, then independent assortment produces $2^N$ genetically different types of gametes and $3^N$ different genotypes amongst the offspring.

The average human has 6700 heterozygous loci carried on 23 pairs of chromosomes. If we were to assume that the 6700 heterozygotic loci are equally distributed along the 23 pairs of chromosomes, and that no crossing over occurs at all at meiosis, a human parent could produce $2^{23} = 8\,388\,608$ genetically different types of gametes, which on fertilization could produce $3^{23}$ different genotypes among the offspring (around $9.4 \times 10^{11}$). The number of genetically different individuals that could arise from fertilization would be greater still. Hence independent assortment can produce huge genetic variation between parents and offspring, variation on which natural selection can act.

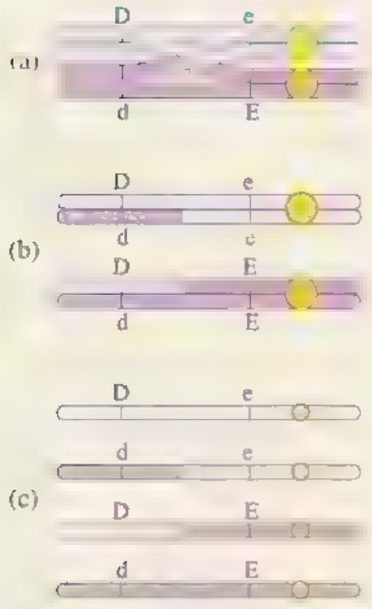



**Figure 3.8** The results of crossing over between a pair of chromosomes (a) at metaphase 1, (b) after anaphase 1, and (c) the chromosomes produced by meiosis.

Independent assortment generates new combinations of alleles on different chromosomes. However, this huge potential for variation is further increased by crossing over between alleles at different loci on the same chromosome. If crossing over takes place between two heterozygous loci on the same pair of homologous chromosomes, it could produce four genetically different kinds of gametes. For example, if one chromosome of a homologous pair carries the alleles *D* and *e* and its homologous pair alleles *d* and *E*, and if there is no crossing over between these alleles, the gametes can have the genotypes *D e* and *d E* only. However, if crossing over does take place between these loci, as in Figure 3.8, the resulting gametes have genotypes *D E* and *d e* also.

Crossing over has produced new allele combinations on chromosomes. If an individual has *N* heterozygous loci, then each could produce $2^N$ genetically different types of gametes which could potentially produce $3^N$ genetically different individuals on fertilization. Therefore with about 6700 heterozygous loci in a human, potentially $2^{6700}$ different genetic combinations are possible in the gametes. You can see that the vast number of possible allele combinations that can arise during recombination through crossing over and independent assortment in organisms with heterozygous loci generates a huge reservoir of genetic variation for natural selection to act upon, which forms the gene pool.

This potential for genetic variation arises through reassortment and recombination at meiosis, followed by syngamy between gametes in sexually reproducing organisms. This potential does not exist for organisms that reproduce asexually. The evolution of sexual reproduction about 1500 Ma ago was followed by a period of rapid protoctist diversification. However, sexual reproduction does not appear to occur in some protoctist lineages. The advantages and disadvantages of sexual and asexual reproduction are explored in Chapter 4, but it is important to note that success, in evolutionary terms, is not confined to organisms that can reproduce sexually.

The original source of new alleles is through the **mutations** that give rise to new DNA sequences. Point mutations are changes in the sequence of base pairs in the DNA due to replication errors or chemical instability in some of the DNA bases. Mutations may also involve larger structural changes in chromosomes or changes in the number of chromosomes. Mutations may occur in any cell, but only those in cells that produce gametes contribute to genetic variation in sexually reproducing populations.

The consequences of mutations range from the sequence coding for a protein being disrupted so that the product is non-functional, to changes in control genes, to loss of sections of the genetic material or its duplication.

A point mutation may or may not have a phenotypic effect, depending on the actual alteration in the base sequence and the function of the coding sequence. Such a mutation may give rise to a serious malfunction and be eliminated, or it

may improve molecular function and be selected. Many mutations are neutral, in that they are neither selected for nor against. However, the environment of a population of organisms may change with time so that such a mutation is no longer selectively neutral, and the variation it provides could confer an advantage and be selected for and so increase its frequency within a population.

Large changes in the genome may take place if chromosomes break naturally, or when exposed, for example, to chemicals or ionizing radiation. Parts of a chromosome may be lost (a deletion), or the sequence of genes, relative to one another, may change if the pieces of chromosome rejoin in a different order. Sections may also be duplicated. In addition, whole chromosomes may be lost or fuse to produce a new karyotype. Study Figure 3.9 which illustrates a number of possible structural changes and the terms used to describe them.

In an inversion, a segment within a chromosome is rotated through 180 degrees. A translocation is a transfer of segments, usually between non-homologous chromosomes.

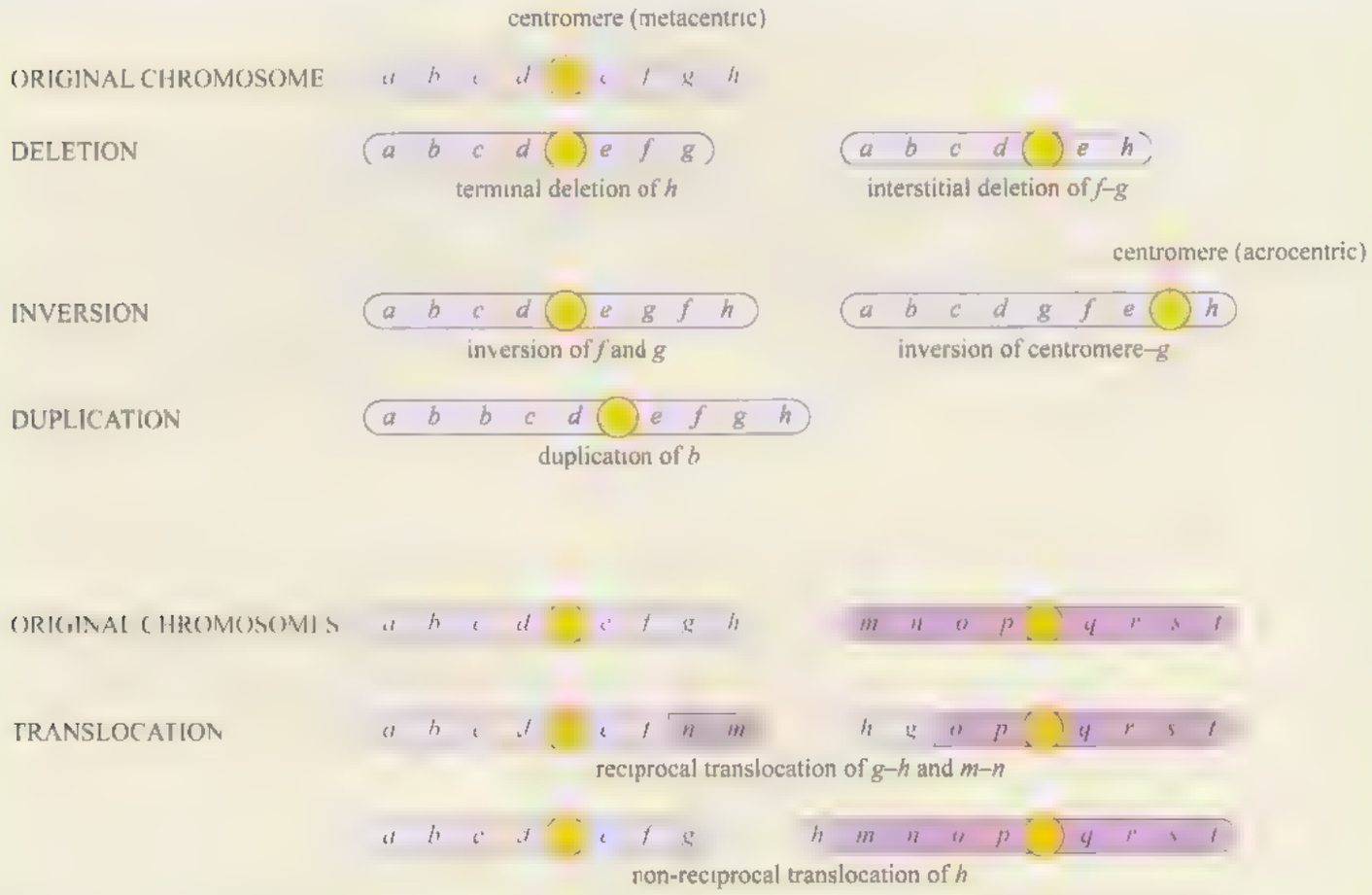



**Figure 3.9** The major types of changes in chromosome structure.

**Figure 3.10** The behaviour of homologous chromosomes during meiosis in an individual heterozygous for an inversion. (a) Pair of homologous chromosomes prior to meiosis. (b) Pairing at prophase I of meiosis. (c) Crossing over and exchange of genetic material. (d) Separation of chromosomes at anaphase I showing (i) a chromatid bridge, with two centromeres; (ii) a fragment without a centromere (both (i) and (ii) give non-viable gametes); (iii) two chromatids with all parts intact (which give viable gametes).

Figure 3.10 shows the irregular behaviour of the chromosomes during meiosis in a cell heterozygous for an inversion.

At prophase I some homologous regions of the chromosomes cannot align linearly, and loops may be formed. Crossing over within the inversion results in some incomplete chromosome fragments, (Figure 3.10d) and half the gametes are defective. Therefore heterozygotes with a large inversion have reduced fertility and hence reduced fitness. However, if the inversion segment is short, the probability of crossing over within the inversion is small and groups of genes within the inversion tend to be conserved. A group of genes that always remain together during subsequent meioses is known as a **supergene**. Such supergenes may conserve adaptations within a chromosome, so that they are not split up by recombination, so increasing the fitness of those individuals that carry it.

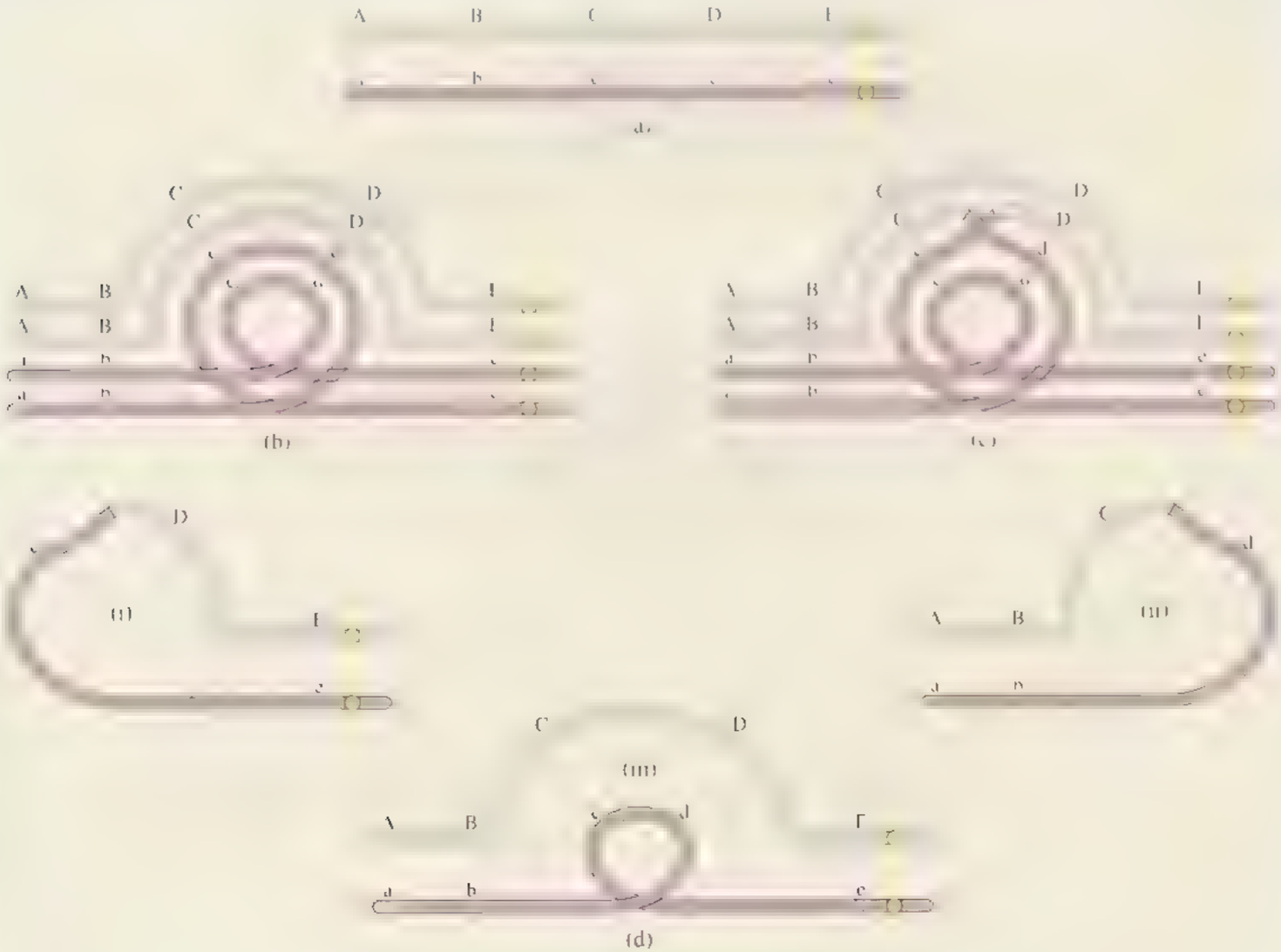



**Figure 3.11** Pin and thrum plants of the primrose, *Primula vulgaris*.

One example of a supergene is found in the reproductive system of the primrose, *Primula vulgaris* (see Figure 3.11). This supergene complex includes loci determining style length, anther height, pollen size, and style and pollen incompatibilities and produces two 'super-alleles', *s* and *S*. Pin plants, with long styles and low anthers, have the genotype *ss* and thrum plants, with short styles and high anthers, have the genotype *Ss*.

The mechanical and physiological differences between the two plant morphs mean that generally, pin and thrum plants cross-pollinate, a mechanism for promoting outbreeding (the fusion of gametes from different individuals). Breeding mechanisms and the genetic consequences of outbreeding and inbreeding (breeding with closely related individuals or selfing) are considered in more detail in Chapter 4.

Now study Figure 3.12 overleaf. The consequence of a translocation can be the positioning of alleles from previously non-homologous chromosomes on the same chromosome. In this figure, half the gametes are defective, so the fertility of the heterozygote is reduced. However, if genes from two non-homologous chromosomes come into close proximity in this way, the probability that they are passed on to subsequent generations as a unit is increased, so adaptive combinations of genes may remain together.

In addition to the chromosome mutations that have been described above, recent studies have shown that mutations, including deletions and inversions, may arise when transposable genetic elements encode transposases, enzymes that enable these transposable elements to 'jump' around chromosomes, altering their structure. Transposable elements appear to exist in all organisms, both prokaryotic and eukaryotic, and can move around the genome and cause mutations, for example when they are inserted into a gene sequence.

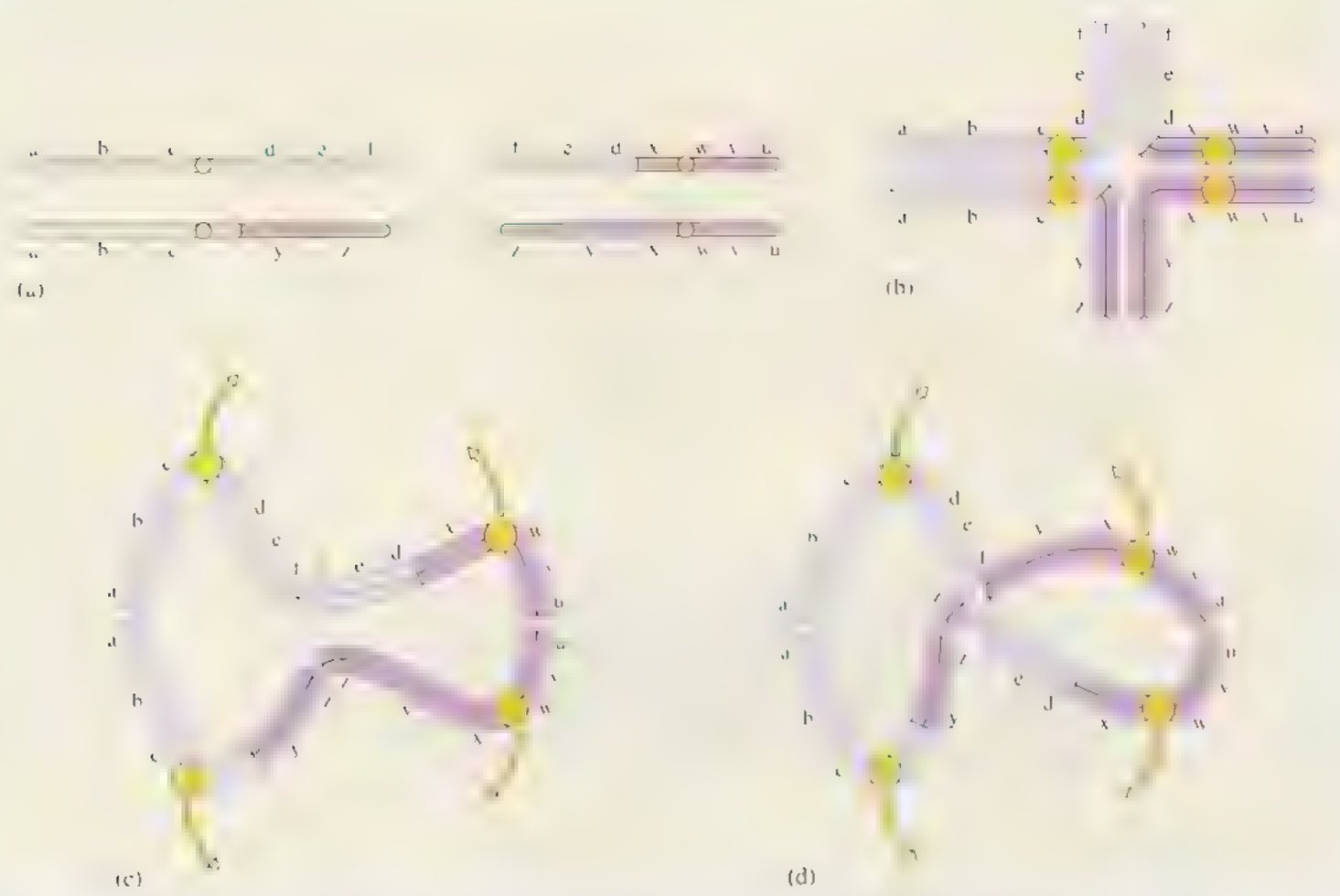


**Figure 3.12** The behaviour of four chromosomes with a reciprocal translocation. (a) Chromosomes with reciprocal translocation. (b) Pairing during meiosis. (c) and (d) Alternative configurations for separation of the chromosomes at anaphase I. (c) gives rise to non-viable gametes because one chromosome segment is duplicated and another is missing in both sets of chromosomes. (d) gives rise to viable gametes as both upper and lower combinations are balanced with all parts of the two chromosomes.

Changes in chromosome numbers may come about as the result of errors in cell division. Individual chromosomes may be lost or gained, resulting in a situation known as **aneuploidy**. An organism with an extra chromosome is trisomic for that chromosome and an organism lacking one member of a pair of chromosomes is monosomic. Trisomics have been described in *Datura stramonium* (Jimson weed or thorn apple). There are 12 chromosomes in the haploid set and trisomics are known for all of them, each affecting the appearance of the seed capsule in a different way and hence generating new phenotypes. Monosomics in *Datura* are non-viable, and it is generally found that having some genes and chromosomes in numbers greater than normal is less deleterious than having them in less than normal quantities.

Let us move on to consider mutations that change the numbers of sets of chromosomes.

› What would be the consequence of failure of cell division after anaphase II of meiosis for any resulting gametes? (You may find it helpful to refer back to Figure 3.5 to remind yourself of the outcome of a normal division II of meiosis.)

- The resulting gametes have double the number of chromosomes.

Note that diploid gametes could also arise following abnormal segregation of chromosomes at anaphase I of meiosis.

Any viable progeny arising from fertilization of such a gamete by a normal haploid gamete are triploid and sterile because their cells contain an extra set of chromosomes that cannot segregate regularly at meiosis, and the sets of three homologous chromosomes are unevenly distributed. Thus any gametes are likely to be lacking some chromosomes and have extras of others. However many plants can reproduce asexually, (which will be described further in Chapter 4), so such a sterile individual may survive. Should a further doubling of chromosomes occur, fertility may be partially restored and a viable, sexually reproducing population (autohexaploid), genetically distinct from the parent population could become established (Figure 3.13 overleaf, route A). Gene flow from the parent population is severed, so a new, **autopolyploid** ('auto' meaning 'self' and 'polyploid' having more than two chromosome sets) species has evolved, sympatric speciation has occurred.

If two diploid gametes fuse, autopolyploid cells with a tetraploid genome are formed (Figure 3.13, route B). Autopolyploidy may also arise following abnormal mitosis during development. Subsequent mitotic division of such cells may result in an autopolyploid individual. Such tetraploids often have low fertility because of complications during meiosis. As with the autohexaploids whose origin is described above, each chromosome has more than one homologous partner with which to pair. Pairing may take place between any number of homologues and is frequently followed by failure of the chromosomes to separate to give equal numbers of chromosomes in each gamete. Autopolyploids are often found to reproduce vegetatively, for example the cultivated potato, *Solanum tuberosum*, which is apparently a tetraploid that has arisen naturally from a wild diploid South American ancestor.

If a hybrid is formed between two non-matching genomes, such as two species A and B, the hybrid is generally unfit because of genetic incompatibilities. Hybrids were made in 1900 at Kew by crossing two commonly cultivated species, *Primula floribunda* and *P. verticillata*. These hybrids had the same chromosome number (18) as each of the parents, and were intermediate in appearance between them, but were sterile. In this hybrid situation, meiosis would be irregular because the chromosomes lack homologues with which to pair. However a very small number of plants were later observed to set seed and were found to have 36 chromosomes; they were tetraploids. These fertile polyploids had arisen by doubling of the hybrid chromosomes and were able to undergo meiosis normally as illustrated in Figure 3.14 overleaf (where only 6 chromosomes per parent species are illustrated for simplicity). Such **allopolyploid** tetraploids (formed from two species) can produce fertile gametes as each chromosome has only one homologue with which to pair, whereas autopolyploids have lower fertility because there are more than two homologues of each chromosome.

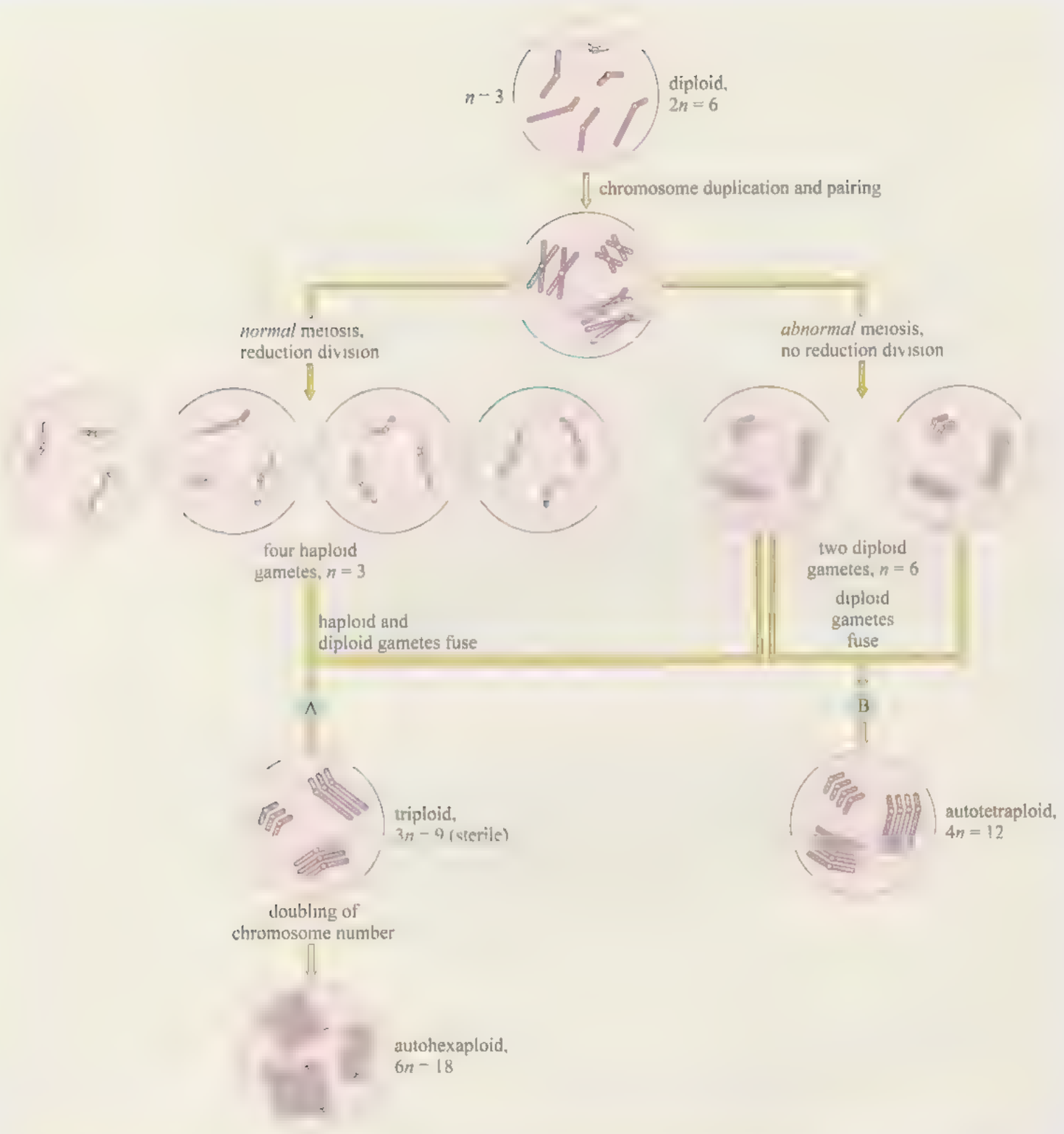



**Figure 3.13** Routes by which autopolyploidy may arise.

Notice that the allotetraploids are reproductively isolated from both parents because of the low fertility of any hybrids between them. They are effectively a new species, known as *P. kewensis*, distinct from both parents and unable to produce fertile seed when crossed with either parent. Hybridization followed by polyploidy could produce a new allopolyploid species in only three generations: this process is another way in which sympatric speciation may take place.

It has been estimated that roughly 50% of flowering plant species have originated as polyploid hybrids. However polyploidy is comparatively rare in animals, probably because it is difficult to establish in sexual, outbreeding species with sex determining mechanisms. Polyploidy is found in some groups where asexual reproduction occurs, such as a very few insects, crustaceans, earthworms, fish, amphibians and reptiles.

Polyploid plants are often found to be larger than their diploid parents, with thicker, fleshier leaves and larger flowers and fruit. The grain yields of durum and bread wheat, *Triticum durum*, a tetraploid, and *Triticum aestivum*, a hexaploid, provide examples of desirable characters in plants cultivated for food.

A polyploid may be better adapted than either parent to a particular environment. The allopolyploid cord grass *Spartina anglica* arose around 1870 in Southampton Water following chromosome doubling of a hybrid between an American species, *S. alterniflora* and the native British species, *S. maritima*. *S. anglica* was found to grow particularly well in mudflats and has been widely planted to stabilize coastal environments. The distribution of both *S. alterniflora* and *S. maritima* has regressed during the last 100 years, but *S. anglica* is widespread around the coast of the British Isles.

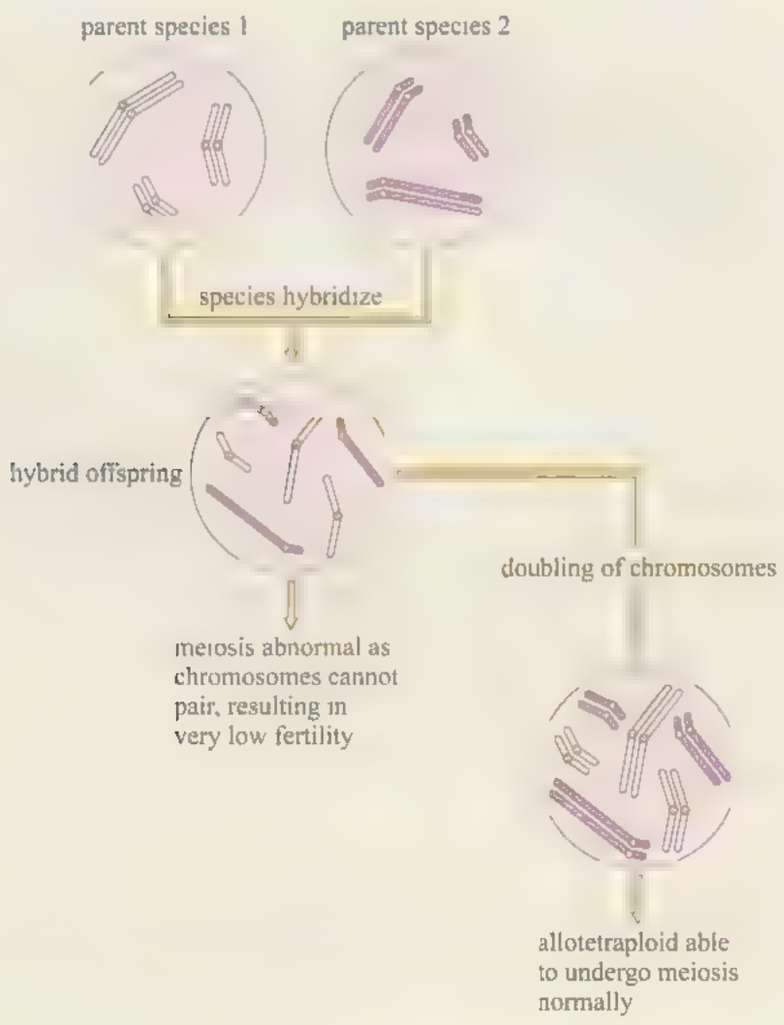



**Figure 3.14** Allopolyploidy following hybridization.

### SUMMARY OF SECTION 3.4

1 Independent assortment and recombination during meiosis, followed by fertilization, can generate huge genetic variability.

2 The source of new alleles is through mutations that give rise to new DNA sequences arising as point mutations or as a result of rearrangements of larger segments of chromosomes.

3 Chromosomal mutation may result in situations in which groups of genes may remain clustered together through meiosis as supergenes, or in polyploidy where the genetic changes may bring about reproductive isolation and speciation within a very small number of generations.

## 3.5 NATURAL SELECTION AND EVOLUTION

Charles Darwin and Alfred Russel Wallace laid the foundations for modern evolutionary thinking in their joint communications on the theory of evolution by natural selection presented in 1858. We began this chapter with the four basic premises of their theory and have examined the origin and expression of genetic variation and its effects on relative fitness when organisms are subject to natural selection. Genetic drift can change the relative frequencies of alleles in the gene pool of a population by chance. Mutation, genetic drift and natural selection may all contribute to the genetic make-up of populations which can change with time and changing environment.

Evolution is generally taken to mean cumulative change, and **biological evolution**, more specifically, the change in characteristics of descendant populations of organisms. These changes in phenotypes in populations over time may accumulate over generations to produce significant alterations and eventually new species. Today, most scientists studying evolution consider that all living organisms ultimately share a common ancestry in simple organisms that arose thousands of millions of years ago. From simple beginnings, evolution has given rise to a huge diversity of organisms, some extremely complicated, but others still unicellular. Therefore biological evolution cannot be viewed only as a process by which more sophisticated organisms replace older, simpler ones. The existence of a species in the present day demonstrates its evolutionary success, regardless of whether it has been highly modified by evolution, or like the 'living fossil' fish, the coelacanth, *Latimeria chalumnae*, it has remained apparently unchanged in morphology for many millions of years.

The process of evolution and its study integrates the ecological interactions between organisms (in the struggle) and their genetic make-ups (variation) and the effects of natural selection on the genetic composition of offspring (principle of inheritance). Recall, however, that natural selection may be stabilizing, and hence may not result in evolution, that mutation is a source of new genetic variations, that polyploidy can lead to speciation and that genetic drift can impact on gene frequencies in populations, resulting in different outcomes from that of natural selection alone.

### SUMMARY OF SECTION 3.5

Evolution may take place through changes in the gene pool as a result of mutation, genetic drift or natural selection, singly or in combination.

## REFERENCE

Tauber C. A. and Tauber M. J. (1977) Sympatric speciation based on allelic changes at three loci: evidence from natural populations in two habitats. *Science*, **197**, pp. 1298 -1299.

# REPRODUCTION

## 4.1 INTRODUCTION

During the course of their lives, animals and plants generally have periods of growth followed by periods of reproduction. One of the most important aspects of an organism's life history is the way it reproduces. There are two main ways in which organisms reproduce: sexually and asexually.

Asexual reproduction occurs naturally in most animal and plant phyla as well as in microbes. Indeed, alternation between sexual and asexual phases of the life cycle is common in plants and in members of some invertebrate taxa. However, obligate or strictly asexual plant and animal 'species' are rare. For example, in animals they comprise little more than one in every 1000 of the named species.

In its broadest sense, **sexual reproduction** can be defined as any process in which genetic material is transferred from one cell to another. Meiosis (reduction division of chromosomes) and syngamy are two essential features of sexual reproduction that are universal for all organisms.

As summarized in the left part of Figure 4.1, during meiosis the parent cell, which is **diploid** (carries two sets of homologous chromosomes), divides, resulting in the production of four daughter cells, each of which is **haploid** (contains only one set of chromosomes). The haploid products of meiosis are known as gametes. Gametes fuse to form a zygote during the process of syngamy (fertilization). In the course of meiosis, **recombination** occurs, as a result of the exchange of genetic material between homologous chromosomes by crossing over followed by the random allocation of one member of each homologous pair to each of the haploid gametes (Section 3.4). The essential feature of meiosis is that it 'shuffles the pack'. Usually, syngamy occurs between gametes derived from different individuals and this process, together with recombination during meiosis, has an important genetic consequence.

❯ What is the genetic outcome of recombination and syngamy?

● Offspring differ from one another and from their parents.

Each offspring is a unique product of fusion of two meiotically diversified haploid nuclei and is therefore different from either of the diploid parents from which it is descended.

In contrast, **asexual reproduction** does not involve meiosis or the production and fusion of gametes. Instead, a single individual produces exact genetic copies of itself through mitosis. These copies are called **clones** (Figure 4.1, right).

❯ What are the genetic consequences of asexual reproduction?

● The parental genome is conserved. Offspring are genetically identical to their parent and to each other, unless mutation occurs.

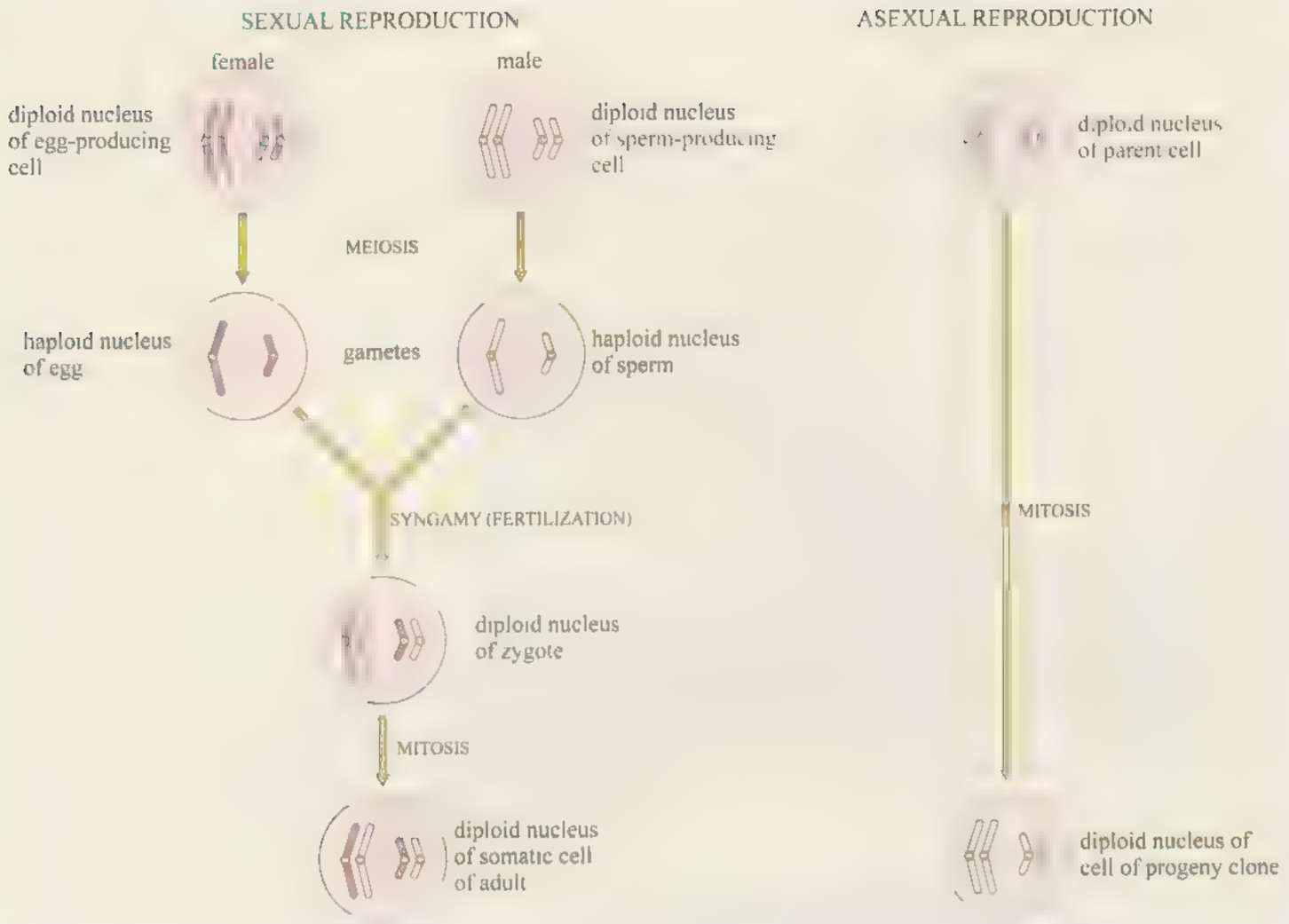



**Figure 4.1** The products of sexual and asexual reproduction. In sexual species (left), meiosis results in the production of four haploid gametes. (For simplicity, only one of each of the four gametes is shown.) Gamete fusion (syngamy or fertilization) results in a diploid zygote, with each parent contributing only half of the genome. The zygote undergoes mitosis to produce individuals that are genetically different from each other and from both parents. Asexual organisms (right), which reproduce by mitosis, produce identical copies of themselves (clones).

Sexual and asexual reproduction can be regarded as two extremes of a continuum. Between these extremes is a host of reproductive modes which incorporate features of both sexual and asexual reproduction.

In this chapter we will explore several aspects of reproduction in eukaryotes, beginning with some examples of the diversity of modes that exist. We will then look at the question of 'why sex?', a question that, even today, puzzles evolutionary biologists. Finally, we will examine some of the consequences of the evolution of sexual reproduction.

## 4.2 ASEXUAL REPRODUCTION

There is little doubt that asexual reproduction evolved long before sexual reproduction and a variety of modes persist today. Several are found in the protoctists. In this section, we examine some of the asexual reproductive modes of animals and plants.

### 4.2.1 VEGETATIVE REPRODUCTION

**Vegetative reproduction** is defined as any mode of reproduction that does not involve the production of eggs. Therefore, in organisms that reproduce in this way, the distinction between growth and reproduction is often blurred. Offspring are derived from a single parent and, because there is no meiosis or syngamy, they are clones of the parent. Several different forms of vegetative reproduction exist.

Many starfish (Echinodermata) and sea-anemones (Cnidaria), as well as undergoing sexual reproduction, sometimes reproduce by an asexual process known as **fission**. Starfish split into two and a new individual is formed from each half. Sometimes an animal can be broken up into two or more pieces, each of which grows into a new individual. This process is known as **fragmentation**, and as a means of reproduction, it depends on the organism having a good regenerative capacity. Sponges (Porifera), hydroids and sea-anemones have astonishing powers of regeneration. For example, if a sponge is macerated by passing it through a fine gauze, the separated cells come together in groups and grow into new individuals. Free-living flatworms (Platyhelminthes) can also reproduce by fragmentation: very small fragments of worm grow into new individuals.

**Budding** is a mode of reproduction that involves the formation of an outgrowth which, on detachment from the parent, develops into a self-supporting individual. It is common in hydras, certain flatworms and several kinds of annelids. In an adult hydra, the dividing cells are found along the column surrounding the enteron. If food is plentiful, more cells form and the extra cells migrate and accumulate at the outer surface of the column, producing a bud. The bud grows its own tentacles while still attached to the parent. Finally it separates from the parent and becomes a free-living hydra (Figure 4.2).

In plants, vegetative reproduction depends largely on the activity of **meristematic cells**, groups of cells (meristems) that are capable of dividing and developing into new tissues. Vegetative reproduction usually takes place via modified stems or roots. In strawberries, for example, a horizontal stem known as a runner grows out from one of the axillary buds on the stem of the parent plant (Figure 4.3). Later the new plant is separated by the withering of the runner.


**Figure 4.2** Asexual reproduction through budding in *Hydra viridis*.

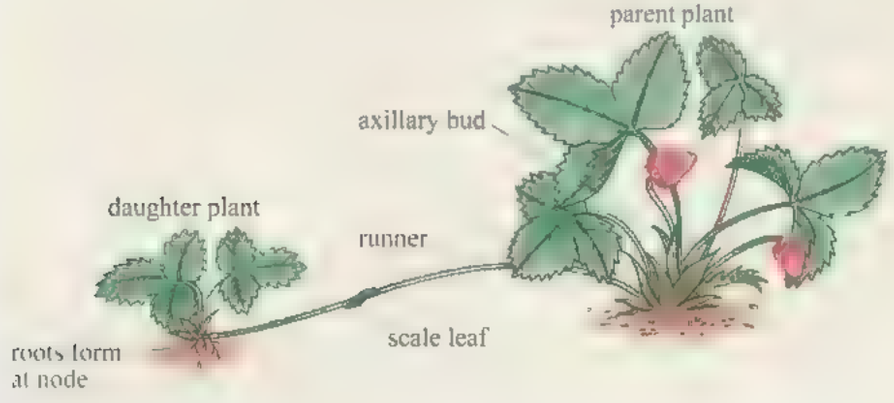



**Figure 4.3** Asexual reproduction in the cultivated strawberry (*Fragaria ananassa*). A horizontal stem, or runner, grows out from the axillary bud and forms roots at the nodes (the 'joints' where leaves arise from the stem).

Vegetative reproduction may also involve the formation of some kind of storage organ which lies in the soil over winter and develops into one or more plants the following year. Such devices are known as **perennating organs** and may be formed from a modified stem, root or bud. For example, garlic (*Allium sativum*) forms a clump of cloves, which are actually underground buds used by the plant for storage. Each clove eventually develops into a new plant, complete with roots (below ground) and shoots (above ground). Another example of a plant that uses this kind of reproduction is the potato (Figure 4.4). A potato is a swollen stem and, like any stem, has both apical buds and axillary buds at leaf nodes. Each 'eye' on the surface of a potato comprises an axillary bud with its scale leaf. It is from a bud that a new plant is formed.

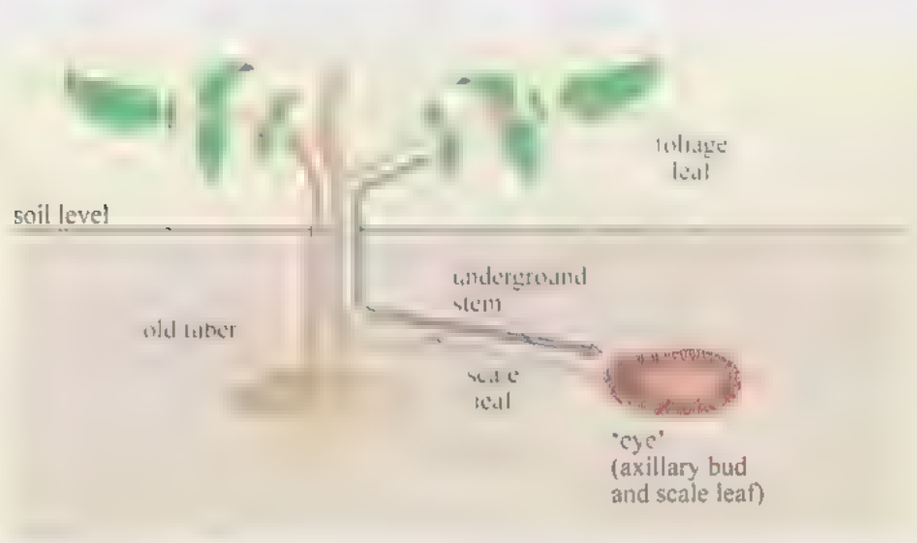



**Figure 4.4** The formation and utilization of perennating organs in the potato plant. Towards the end of the growing season, the plant forms swollen tubers at the ends of horizontally growing underground stems, using food materials transported into it from the above-ground foliage leaves (blue arrow). The tubers lie dormant in the soil until the following year when axillary buds give rise to new plants. The old tuber shrivels as the food stores within it are solubilized and transported out to the regions of growth (red arrow).

Vegetative propagation from root systems also occurs in plants. The plant sends out adventitious roots (i.e. roots produced from a stem) which then develop into separate shoot systems. The creosote bush (*Larrea tridentata*) of the Mojave desert in California propagates in this way. The daughter plants form rings around the original parent, which expand in diameter at a rate of about one metre per 500 years. From the size of the rings, the Mojave desert bushes are estimated to have been propagating via the adventitious budding of root systems for up to 12 000 years.

### 4.2.2 PARTHENOGENESIS

Parthenogenesis is traditionally viewed as being an asexual process and sometimes the terms parthenogenesis and asexual reproduction are used as synonyms. However, classifying all parthenogens as asexual is confusing because, in some types of parthenogenesis, meiosis and syngamy take place.

- Given that there is meiosis and syngamy in some parthenogens, why would equating parthenogenesis with asexual reproduction be confusing?

- Because meiosis and syngamy are regarded as sexual processes.

**Parthenogenesis** literally means 'virgin descent'. In obligate parthenogens, there are no males and new individuals are produced from a single female parent. Several different forms of parthenogenesis have been identified, which fall into two main categories: **arrhenotoky** (pronounced 'ah renn oh toe-key', and derived from the Greek, meaning 'male-begetting'), which is the production of haploid males from unfertilized eggs, and **thelytoky** (pronounced 'thelly-toe-key', literally 'female-begetting'), the production of diploid females from unfertilized eggs.

Arrhenotoky is common among bees, wasps and ants (order Hymenoptera), and certain beetles (Coleoptera) and bugs (Hemiptera). Females produce haploid eggs, via meiosis, which develop into diploid females if fertilized, and into haploid males, if not (Figure 4.5). In honeybees, for example, the queen is the only sexually mature female in the colony. She receives only one insemination with sperm during her life. She stores the sperm and for several years produces eggs which she can either fertilize with stored sperm as they are laid, or not. Fertilized eggs develop into diploid females, most of which are sterile workers, but a select few become fertile queens. Unfertilized eggs develop into haploid male drones.

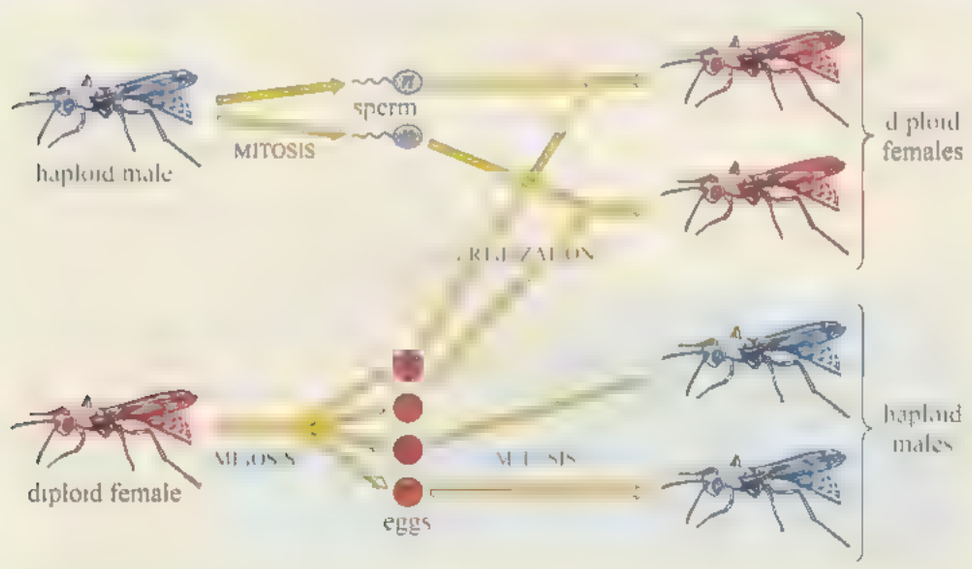



**Figure 4.5** Arrhenotoky involves the parthenogenetic production of haploid males from unfertilized eggs (shown in blue shaded area) and occurs in some species of insects. Diploid females are produced through fertilization (i.e. sexually).

Thelytoky can take several different forms. In one form, each diploid female develops from a single, mitotically produced cell. There is no meiosis or syngamy. Mothers produce daughters that inherit their complete, unrecombined diploid genome. This kind of reproductive mode is called **apomixis** (also known as ameiotic parthenogenesis) and is widely distributed in both animals and plants. It is known to occur in bees and wasps (Hymenoptera), and certain bugs (Hemiptera), flies (Diptera), grasshoppers (Orthoptera), lizards (Reptilia), frogs (Amphibia) and in several crustaceans (Arthropoda), flatworms (Platyhelminthes) and roundworms (Nematoda). Figure 4.6 illustrates apomixis in the whiptail lizard (*Cnemidophorus tesselatus*).

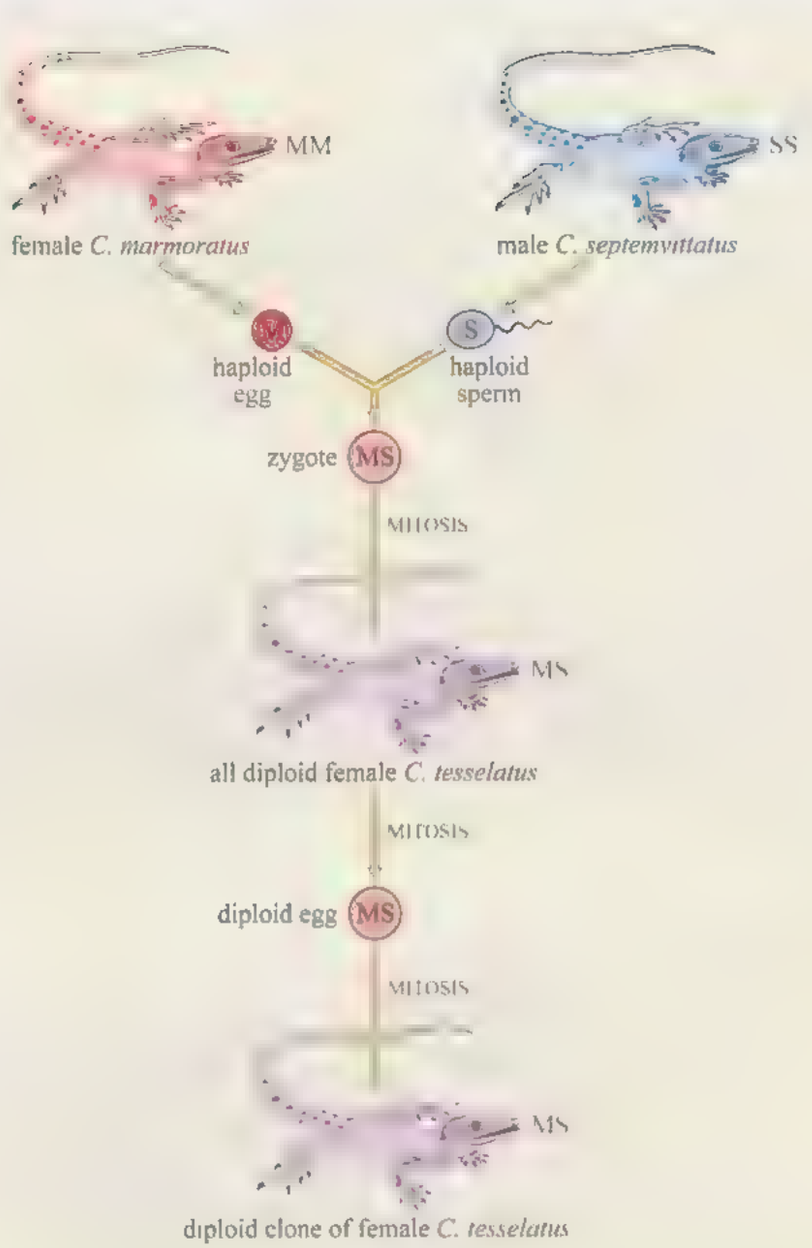



**Figure 4.6** Apomixis, the production of diploid females from diploid eggs in the whiptail lizard (*Cnemidophorus tesselatus*). Diploid unisexual (all female) *C. tesselatus* (genome MS) are derived by hybridization between two bisexual (sexually reproducing) lizard species, *C. marmoratus* (female, MM) and *C. septemvittatus* (male, SS). The diploid *C. tesselatus* females produce diploid eggs (by mitosis), in which chromosomal recombination has not occurred. These eggs develop, without sperm, into diploid offspring which are clonal, i.e. genetically identical to each other and to their mother. This parthenogenetic mode of reproduction can perpetuate clonal lineages indefinitely, yielding genetically identical females, generation after generation.

In flowering plants, apomixis is the production of seeds without normal meiosis or fertilization taking place; the seeds are generated by mitosis. For example, in citrus fruits and dandelions, an embryo develops from a diploid cell in the ovule that has not undergone meiosis. As the embryo develops, the surrounding tissues form the seed and fruit in the normal way. One advantage of apomixis in plants over vegetative reproduction is that it enables the plant to spread its offspring over a large area using the methods of seed dispersal typically available to embryos formed by sexual reproduction. Dandelions, for example, are able to disperse their seeds as far as the wind can take them.

) In apomicts, to what extent are offspring genetically different from their parent?

● They aren't. Neither normal meiosis nor syngamy occur, so offspring are genetically identical to their parent.

A second form of thelytoky is known as **automixis** (or meiotic parthenogenesis). In automicts, females produce diploid eggs via meiosis. There are no males, so meiosis is not followed by fertilization and the consequent doubling of chromosome number. Instead, automictic females have an internal solution for doubling the chromosome number and so producing diploid eggs, which then go on to develop by mitotic division into adults. A number of chromosome doubling devices have been described. One of the most common is **premeiotic endomitosis**, in which is mitosis without cell division takes place before meiosis. This process is shown in Figure 4.7.

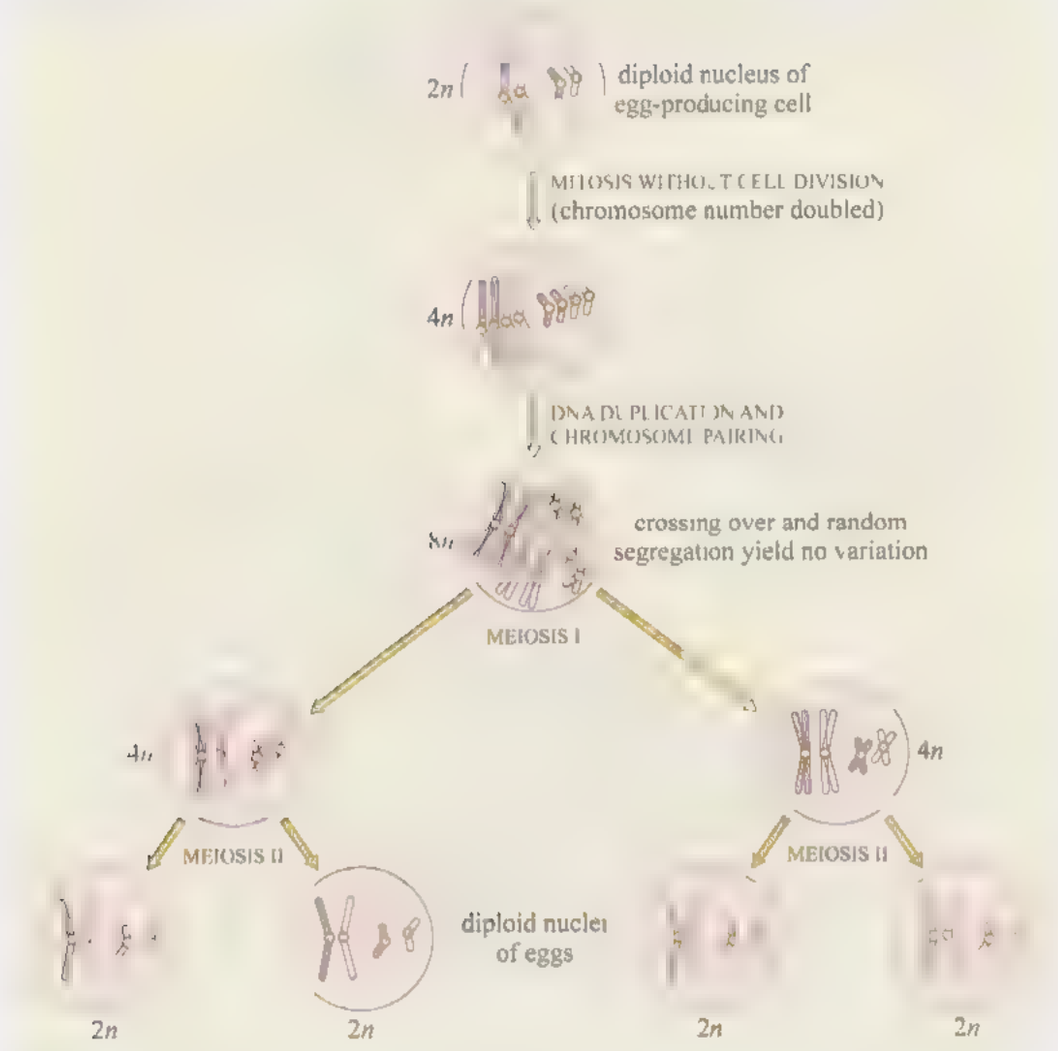



**Figure 4.7** Production of diploid eggs by premeiotic endomitosis followed by meiotic chromosome duplication and cell division.

- What is the effect of premeiotic endomitosis on chromosome number?

- It doubles the number of chromosomes in the egg before meiosis begins. Thus at the start of meiosis the cell is $4n$, so after the chromosome doubling of meiosis and just before division I, the cell is $8n$.

At the end of meiosis II, therefore, each egg contains the diploid number of chromosomes. This mechanism of producing diploid eggs is found in the lizard *Cnemidophorus uniparens* and in at least two species of fish (genus *Poeciliopsis*, top minnows). It is also known to occur in grasshoppers (e.g. *Moraba virgo*) and in parthenogenetic earthworms (Annelida).

❍ What are the genetic consequences of the system shown in Figure 4.7?

● It generates no variation. Crossing over happens only between identical chromatids (since the chromosomes are duplicated), so generates no variation. Pairs of chromosomes from different homologues in the parental diploid cell segregate, but since they are identical, no variation is introduced into the diploid eggs produced. In effect, the result is the same as that for apomixis, i.e. the egg is genetically identical to the somatic cells of the mother.

In addition to the asexual systems described above, there are several others that are difficult to classify as either sexual or asexual. One of the most common of these modes is known as **hybridogenesis**. Unlike thelytoky, in hybridogenesis males are required for successful reproduction. The uniqueness of the system lies in the fact that the males are of a different, but closely related, species. These males (species 1) mate with the females (species 2) to form a diploid hybrid female. This hybridogen (the result of a cross between two different species) then mates with a male of species 1 but during meiosis, the male's genome is excluded so that all the eggs carry only the unrecombined genome of the mother.

A classic example of fish that reproduce via hybridogenesis is seen in the genus *Poeciliopsis* (Figure 4.8). Here, crosses between two bisexual species, *P. monacha* and *P. lucida* produce an all-female hybrid called *P. monacha-lucida*, which reproduces via hybridogenesis, passing on the maternal genome unchanged through successive generations, while the paternal genome is replaced with each generation.

A second system in which males of a different species are required for successful reproduction is known as **gynogenesis**. Sperm is required to begin embryogenesis (development of the egg into an embryo) but no syngamy between egg and sperm occurs. Gynogenesis occurs in the Amazon molly, *Poecilia formosa* (Figure 4.9). *P. formosa* is derived from hybridization between the bisexual species *P. latipinna* and *P. mexicana*. *P. formosa* females produce diploid eggs via premeiotic endomitosis, so they have not undergone recombination. However, in contrast to apomixis, these eggs require sperm (from a related species) in order to develop. The sperm presumably penetrates the egg but syngamy between egg and sperm nuclei does not occur, so the sperm makes no genetic contribution to the offspring.


**Figure 4.8** Hybridogenesis in fish. Crosses between females of the bisexual species *Poeciliopsis monacha* (containing two haploid M genomes, MM) and male *P. lucida* (containing two haploid L genomes, LL) produce a unisexual female *P. monacha-lucida* (ML). Each *P. monacha-lucida* female generated by the original hybrid cross produces haploid eggs containing only the unrecombined haploid *P. monacha* genome (M) from their mother; the *P. lucida* genome (L) of the father is lost. These *P. monacha-lucida* females then mate with other *P. lucida* males (L′L′), syngamy of egg and sperm occurs, and female offspring are produced (ML′) which carry the clonally inherited *P. monacha* genome (M) from their mother and the sexually inherited *P. lucida* genome (L′) from their father. In this way, the *P. monacha* genome (M) is passed on unchanged through many generations, each time combined with a new and different *P. lucida* genome.

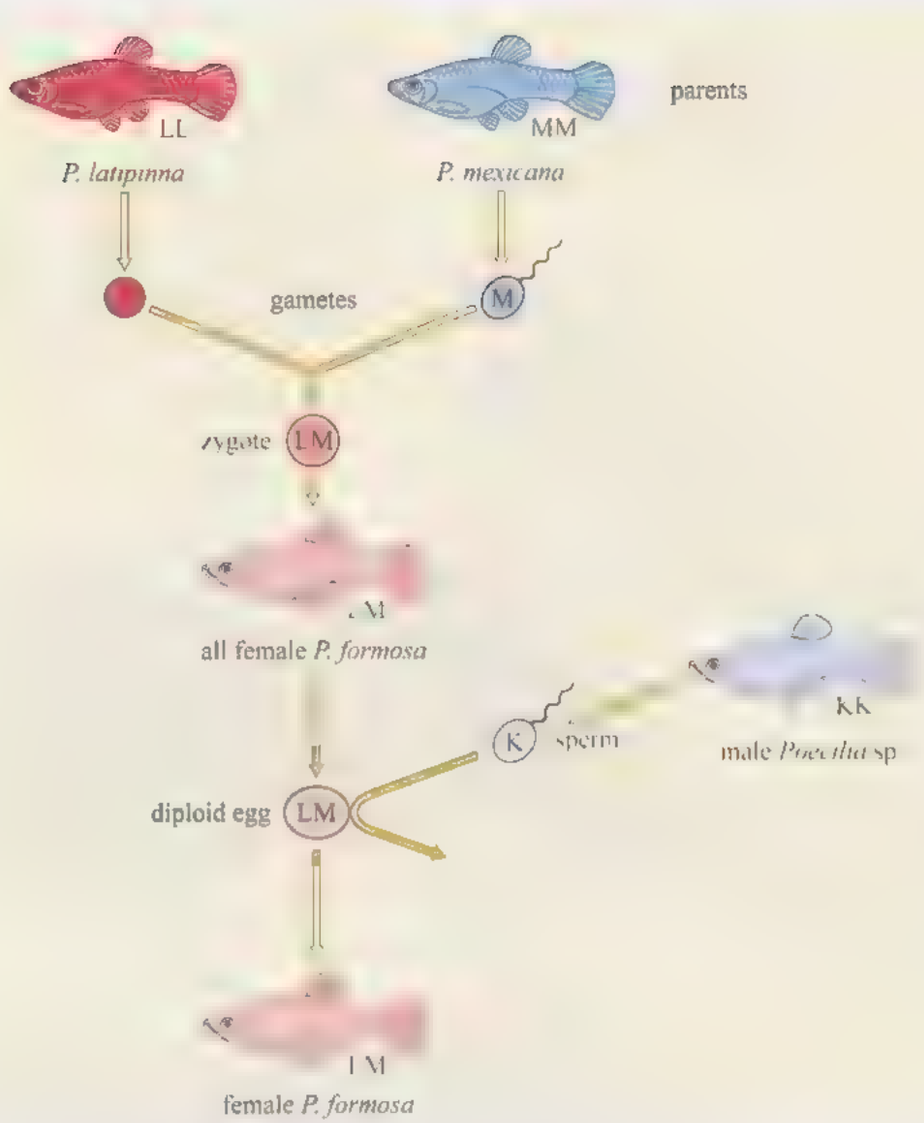



**Figure 4.9** Gynogenesis in the Amazon molly, *Poecilia formosa*. Crosses between the bisexual species *P. latipinna* (genome LL) and *P. mexicana* (genome MM) produce the unisexual female fish, *P. formosa* (genome LM). *P. formosa* produce diploid eggs without recombination. These fish mate with males of a related bisexual species and their sperm stimulates embryogenesis, but syngamy of egg and sperm does not occur. The resulting offspring (genome LM) are genetically identical to their mother and to each other and form a diploid *P. formosa* clone. (In 1% of cases, syngamy between the egg and sperm does occur, resulting in a triploid.)

The three kinds of unisexual modes of reproduction (i.e. reproductive modes that produce only females) discussed above are compared to sexual reproduction in Figure 4.10 and in the following text.

In *apomixis* the hybrid genome (one haploid genome from an ancestral bisexual species A and another from an ancestral bisexual species B) is transmitted to the egg complete, without recombination. The egg develops without sperm into an offspring identical to its mother.

In *hybridogenesis* one ancestral genome (A) from a unisexual female (AB) is transmitted to the egg without recombination, while the other ancestral genome from the male ancestor (B) is discarded; sperm B′ from another male of the same ancestral species then fertilizes the egg, producing an offspring with the original maternal (A) genome and a new paternal (B′) genome.

In *gynogenesis* diploid eggs are produced by either apomixis or premeiotic endomitosis. Sperm from a related bisexual male is then required to stimulate development of the diploid egg, but makes no genetic contribution.

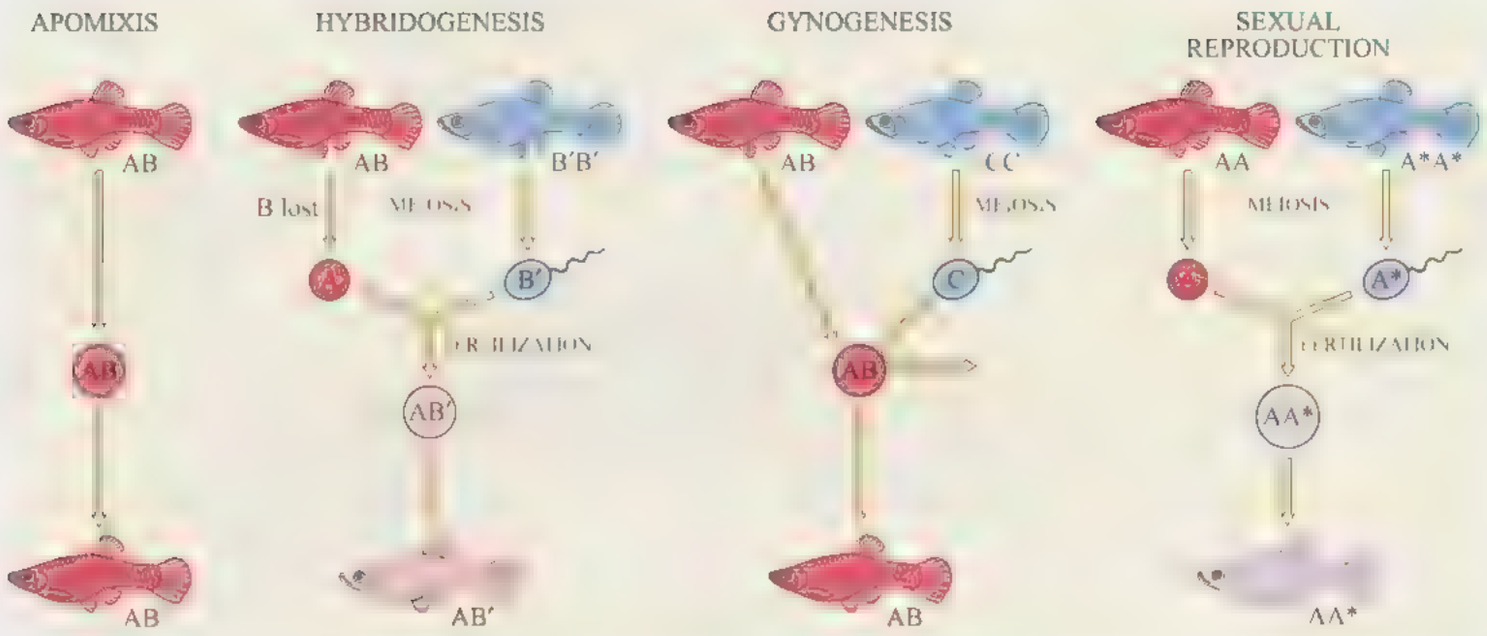



**Figure 4.10** Comparison between unisexual modes of reproduction and sexual reproduction. See text for explanation.

In *sexual reproduction* the mother produces a haploid egg (A) which, due to recombination, contains a unique combination of genes from *her* mother and father; the father produces haploid sperm (A′) in the same way, resulting in offspring that are highly variable.

Unisexual vertebrates of the sort discussed above are most commonly derived from a hybridization event between two species. As a consequence, their establishment in nature is subject to severe genetic, developmental and ecological constraints. They are believed by many who study them to be evolutionary dead-ends. Nonetheless, biologists investigate these animals for the same reasons that medical scientists probe disorders and diseases to define the limits of human health. Identifying the conditions in which asexual lineages prosper or fail provides a window through which biologists can view the adaptive significance of genetic diversity and sex.

## 4.3 INBREEDING AND OUTBREEDING

In the introduction to this chapter, we stated that sexual reproduction usually involves the fusion of gametes derived from different individuals of the same species. If these individuals are very distantly related, the process is known as **outbreeding** (cross-fertilization, outcrossing). Much of what is dealt with in the remaining part of this chapter refers to outbreeding sex. However, it is important to understand that sexual reproduction can, in some cases, involve the fusion of gametes derived from closely related individuals of the same species, a process known as **inbreeding**. The significance of the distinction between inbreeding and outbreeding lies in their genetic consequences.

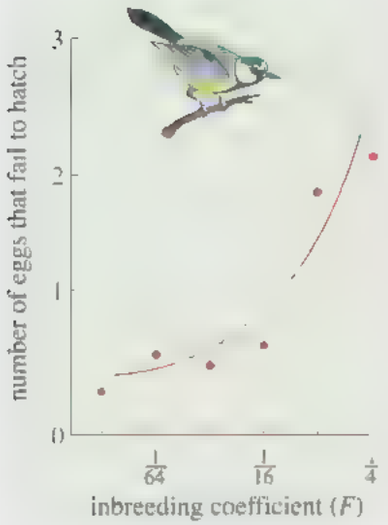



**Figure 4.11** The effect of inbreeding on egg failure in great tits. A measure of the extent of inbreeding is given by an inbreeding coefficient, *F*. The higher the value of *F*, the higher the degree of inbreeding. As *F* increases, so does the number of eggs that fail to hatch. Data from Van Noordwijk and Scharloo (1981).

It is common knowledge that inbreeding can have genetically harmful effects. In the course of history, various families, especially royalty and aristocracy, have practised incest (mating between siblings) or frequent mating between cousins, with dire consequences. Inbreeding in humans leads to premature mortality, mental retardation, albinism and other genetic, physical and mental abnormalities in offspring. The decrease in offspring fitness as a consequence of inbreeding is known as **inbreeding depression**. In genetic terms, its explanation lies in the presence of deleterious recessive mutations in populations. Such genes are not normally expressed, because they are in the heterozygous condition. Due to their common ancestry, close relatives are more likely to carry the same deleterious gene or genes than unrelated individuals. Offspring produced by a union between two closely related individuals therefore have a high probability of inheriting two copies of a deleterious gene or genes (one copy from each parent). Deleterious recessive genes are expressed in these offspring because they are homozygous, but they are not expressed in their parents (who are heterozygous).

Long-term studies on marked populations of the great tit (*Parus major*) have shown that inbreeding depression can have strong effects on reproductive success (Figure 4.11). In song sparrows (*Zonotrichia melodia*), inbred individuals were found to be much more likely to die during population crashes than outbred individuals (Figure 4.12). Perhaps the most powerful studies of inbreeding depression have been done on flowering plants. From these studies, two general observations were made. Firstly, inbreeding effects are easier to detect when plants undergo some sort of environmental stress and secondly, they are more likely to show up later in life (Figure 4.13).

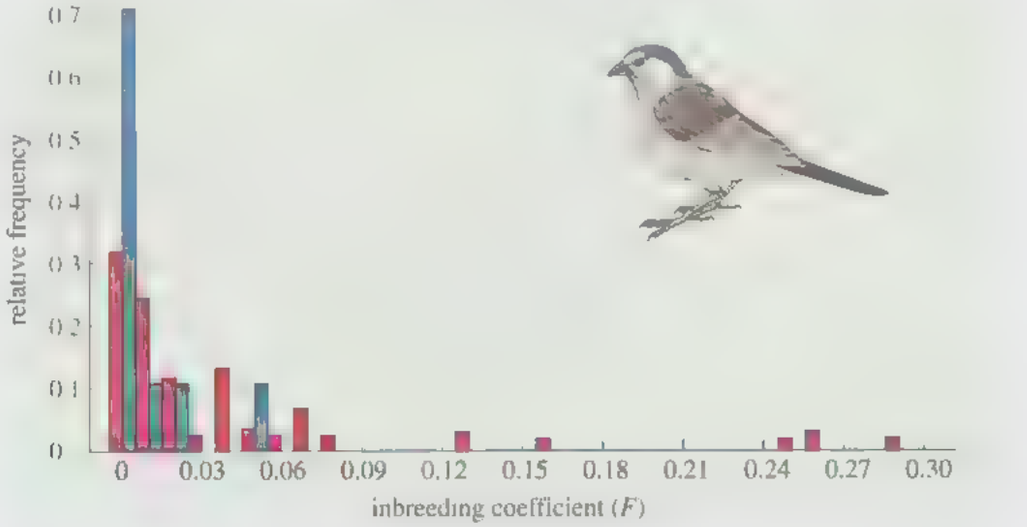



**Figure 4.12** Inbred song sparrows (*Z. melodia*) are less likely to survive a population crash. Several periods of difficult environmental conditions were observed, during which many of the individuals in their population died. The red bars show the proportion (relative frequency) of birds that died with a given value of *F*, the blue bars the proportion that survived with a given value of *F*. The figure shows that 70% of the survivors had an *F* value of 0 (were not inbred), while only 32% of non-survivors had an *F* value of 0. No birds with an *F* value of greater than 0.05 survived. Data from Keller *et al.* (1994).

What can you conclude from Figure 4.13?

As plants get older, the disparity in flower production between outcrossed and self-pollinated plants increases, which suggests that inbreeding depression (the expression of deleterious genes) becomes more pronounced with age.


**Figure 4.13** Comparison of the number of flowers produced (a measure of fitness) as a function of time for outcrossed versus self-pollinated individuals in *Lobelia cardinalis*, a perennial of the harebell family (Campanulaceae). Data from Johnson (1992).

Because of the effects of inbreeding depression, many animals and plants have evolved mechanisms that reduce its occurrence.

There are two main ways in which animals reduce the likelihood of mating with a close relative. The first is by dispersing before sexual maturity, so that close relatives are unlikely to be encountered. Male dispersal is common among mammals, whereas among birds it is more common for females to disperse. A second way is to recognize and avoid mating with close kin, which is often achieved by not mating with individuals with which an individual is reared. There is evidence, however, that for a variety of animals, including mammals, amphibians and some arthropods, kin recognition can occur without individuals having prior experience of each other.

Most plants can both self-pollinate and export pollen to the stigmas of other plants and in many such species, some self-pollination occurs. Mechanisms that decrease the incidence of self-fertilization (selfing), which is an extreme form of inbreeding, include the following:

1 **Dioecy** (pronounced 'die-ee-see'), where individual plants bear either male or female flowers, but not both.

2 In species that are **monoecious** ('mon-ee-shuss'), where male and female flowers occur on the same individual, and in those in which male and female functions occur in the same flower, the likelihood of self-fertilization may be reduced if the anthers and stigma ripen at different times or if the anthers and stigma are separated in such a way that pollen cannot reach the stigma of the flower in which it is produced.

3 Flowering plants may also prevent self-fertilization by being genetically self-incompatible: the pollen from a given flower fails to grow down the stigma of that flower or, less commonly, the embryos resulting from self-fertilization fail to develop.

Although the available evidence suggests that inbreeding does cause a reduction in fitness and that both animals and plants have evolved mechanisms to minimize its occurrence, it cannot be assumed that outbreeding is necessarily more adaptive. As mentioned already, mating with close relatives may result in deleterious combinations of alleles being expressed in offspring. However, inbreeding may also result in the conservation of advantageous combinations of alleles in the progeny, whereas outbreeding tends to break these up. This consideration has led to the idea that there may be an optimal level of outbreeding: very close relatives and totally unrelated individuals should be avoided as mates, but intermediate relatives should be preferred.

Among animals, there is some evidence for optimal outbreeding. For example, laboratory studies on the Japanese quail (*Coturnix coturnix japonica*) show that individuals prefer to associate with first cousins rather than either their siblings or unrelated birds.

In plants, inbreeding may initially result in a decline in mean fitness of the population. However, deleterious recessive alleles that have been masked in heterozygotes by dominant alleles become exposed to and eliminated by selection. In some cases, this process may result in an inbred population with a mean fitness equal to or even exceeding that of the initial outbreeding population. Factors that favour the evolution of exclusive selfing will be discussed further in Section 4.6.

## SUMMARY OF SECTIONS 4.2 AND 4.3

1 Sexual reproduction has two essential features that distinguish it from asexual reproduction: meiosis and syngamy. Meiosis is the process of cell division that is intrinsic to the production of haploid gametes; syngamy is the fusion of two gametes to form a zygote.

2 In sexual reproduction, syngamy can occur between gametes derived from two different individuals that are unrelated, a process known as outbreeding, or between gametes derived from closely related individuals (inbreeding).

3 Asexual reproduction is reproduction without sex, and refers to the production by a single parent of diploid progeny that are exact genetic replicas of that parent. Wholly asexual organisms have no meiosis at all in their life cycles and reproduce by mitotically derived somatic tissues (vegetative reproduction) or mitotically derived single cells (apomixis).

4 Parthenogenesis is the development of a new individual, either male or female, from an unfertilized egg. In some parthenogenetic systems, diploid eggs are produced by one of several chromosome doubling devices.

5 Hybridogenesis has some of the features of sexual reproduction. In a diploid hybridogen, the genome derived from one parental species is transmitted to the egg without recombination, while the genome of the other parental species is discarded. The haploid egg is then fertilized by sperm of a second male of the same species, restoring the hybrid condition.

6 Gynogenetic fish result from crosses between two sexually reproducing species. The resultant hybrid female produces diploid eggs, via either premeiotic endomitosis or apomixis, and then mates with a related bisexual male. The sperm stimulates embryogenesis, but syngamy between egg and sperm does not occur and only the mother's genome is passed on to the offspring.

7 Inbreeding depression usually results from the exposure of deleterious recessive alleles to selection and both plants and animals have evolved mechanisms to reduce its occurrence.

8 In spite of the disadvantages of inbreeding depression, both plants and animals may benefit from some degree of inbreeding.

## 4.4 THE COSTS AND BENEFITS OF SEXUAL REPRODUCTION

By now, you will have realized that many animals and plants reproduce asexually. However, it is also true that many other organisms, including most vertebrates, insects and flowering plants must reproduce sexually or not at all. Although we are all familiar with the widespread occurrence of sexual reproduction, few of us stop to think about why it evolved in the first place. This question may seem an odd one to ask but it has puzzled evolutionary biologists for decades, because there are a number of reasons why sex should not have evolved. It is much simpler and more efficient for animals and plants to reproduce asexually. Why should two parents do the work of one? In order to understand why sexual reproduction persists in nature when a simpler alternative is available, we must identify and quantify, as far as is possible, the relative costs and benefits of sexual and asexual modes of reproduction.

### 4.4.1 THE COSTS OF SEXUAL REPRODUCTION

There are two main costs incurred in sexual reproduction: the cost of recombination and the cost of producing sons, which is also referred to as the cost of meiosis.

#### THE COST OF RECOMBINATION

The process of sexual reproduction involves the rearrangement of nuclear genomes.

❯ What are the three cellular mechanisms of genome recombination?

● (i) Crossing over during meiosis, (ii) random segregation of chromosomes during meiosis and the formation of haploid gametes and (iii) the fusion of gametes during fertilization.

The effect of this 'reshuffling' is to break up genotypes and reassemble them in different arrangements, which can have two consequences. The first is that valuable combinations of particular alleles cannot be maintained by this process. The second is that deleterious combinations of alleles may arise during recombination.

One well-known case of this second effect is the human allele ($Hb^S$) which confers protection against malaria (Figure 4.14). Individuals that are heterozygous ($Hb^S Hb^A$), carrying one allele that confers protection and one that does not ($Hb^A$), are protected against malaria. However, individuals that carry two copies of the $Hb^S$ allele (homozygous, $Hb^S Hb^S$) develop a sometimes fatal condition known as sickle-cell disease and individuals that carry no copies of the $Hb^S$ allele (homozygous, $Hb^A Hb^A$) are susceptible to malaria. When two heterozygous individuals reproduce (sexually), one-quarter of their offspring are likely to succumb to sickle-cell disease ($Hb^S Hb^S$), one-quarter are susceptible to malaria ($Hb^A Hb^A$) and only half of them enjoy the benefits of heterozygosity ($Hb^S Hb^A$).

› Why would it be advantageous if a heterozygous ($Hb^S Hb^A$) mother, living in an area where malaria is common, could reproduce asexually?

● Because all her offspring would be heterozygotes ($Hb^S Hb^A$) and therefore adapted to their environment.

The recombination which occurs in sexual reproduction therefore can extract a cost.

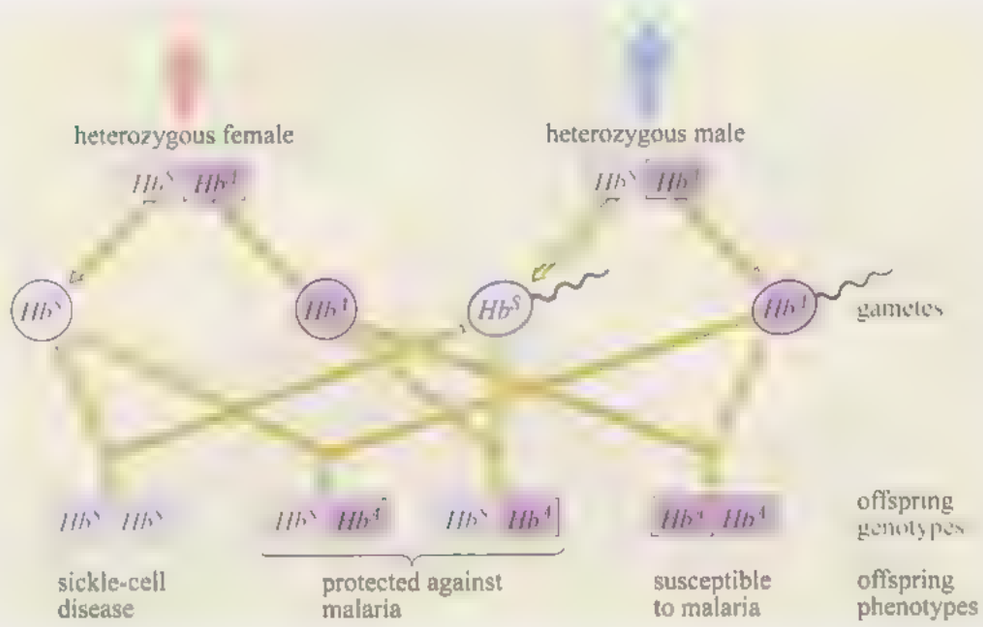



**Figure 4.14** The inheritance of sickle-cell disease. Each heterozygous parent produces two types of gamete, one containing the $Hb^S$ allele and the other carrying the $Hb^A$ allele. When gametes combine at fertilization, three types of offspring result: half of them are protected against malaria (genotype $Hb^S Hb^A$), one-quarter are susceptible to malaria ($Hb^A Hb^A$) and one-quarter have sickle-cell disease ($Hb^S Hb^S$).

### THE COST OF PRODUCING SONS/COST OF MEIOSIS

To understand why sexual females incur a cost by producing sons, we need to compare the reproduction over time, of a sexual female and an asexual female. Imagine a population founded by three individuals: a sexual female, a male and an asexual female (Figure 4.15). At this point, each individual constitutes one-third of the group. Now imagine that each female produces two offspring, after which the parents die. The sexual female produces, on average, one daughter and one son (as explained in Section 4.5.2) and the asexual female produces two daughters (since both are genetically identical to their mother and must therefore be female too). These offspring survive equally well, produce the same number of offspring themselves and their offspring have an equal chance of survival.

**Figure 4.15** The cost of producing sons. F = female, M = male individuals.

| Generation | Sexual reproduction | Asexual reproduction | Fraction of asexual individuals |
|---|---|---|---|
| | F M | F | $\frac{1}{3}$ |
| 1 | F M | F F | $\frac{1}{2}$ |
| 2 | F M | F F F F | $\frac{2}{3}$ |
| 3 | F M | F F F F F F F F | $\frac{4}{5}$ |

From Figure 4.15 you can see that, after one generation, the sexual group consists of two individuals, one male and one female, and the asexual group consists of two females. In other words, there are four individuals in the descendent generation, of which one-quarter are males, one-quarter are sexual females and one-half are asexual females. The proportion of asexual individuals in the population has therefore risen from one-third to one-half in one generation.

› After a further generation, how many individuals will there be in the population?

● Six: four asexual females, since each asexual mother produces two daughters, and two sexual individuals, since each sexual female produces two offspring, one female and one male.

› What fraction of the second-generation population are asexual females?

● Two-thirds.

From these calculations, you can see that it would not be long before asexual reproduction had completely taken over. Maynard-Smith presented this argument algebraically and concluded that the clone of offspring from an asexual female would multiply at twice the rate of progeny descended from a sexual female, due to the cost to the sexual female of producing sons.

This twofold cost of sexual reproduction is sometimes referred to as the 'cost of meiosis', because the sexual female passes only 50% of her genes on to the next generation compared with the asexual female, which transmits 100% of her genes (Figure 4.16). Once again, imagine two females, one sexual and the other asexual (parthenogenetic), both of which leave two offspring. The asexual female transmits two entire copies of her genome to the next generation via her two daughters. The sexual female, on the other hand, transmits half her genome via her daughter and half her genome via her son (the other half being supplied by the father of her offspring).

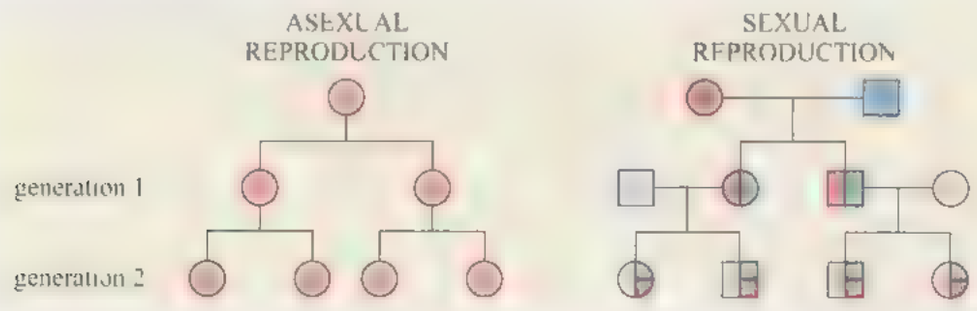



**Figure 4.16** The cost of meiosis. See text for details. The circles denote females and the squares denote males.

› How many copies of the sexual mother's genome are represented in the first generation?

● One copy (a half in each of her two offspring).

In the next generation, each of the asexual female's daughters themselves produce two daughters.

❯ How many copies of their grandmother's genome are represented in the second generation?

● Four copies, one in each of the four grandchildren.

Now look again at the pedigree for the sexual female. Her daughter, carrying half of her genome, produces one daughter and one son, as does her son. In other words, she has four grandchildren.

❯ How many copies of her genome are represented in the second generation?

● One copy — each grandchild has a quarter of her genome.

In one generation, the parthenogenetic female therefore replicates her genes at twice the rate of the sexual female, all else being equal. The sexual female loses out by sacrificing half her genome in meiosis, in comparison to the parthenogen, which transmits her entire genome. In addition, the sexual female produces a large egg that must contain the cytoplasm and nutrient reserves required by the zygote, because the sperm of the male contributes little of these materials. Consequently, the sexual female is providing twice as much cytoplasm and nutrients per genome copy compared to the parthenogenetic female, and suffers a 50% cost in the efficiency of genome propagation as a result. This paradox raises the important question of why a female should raise an offspring that carries only half her genes when *for the same effort* she could raise one that carries all her genes. We will come back to this topic later.

In addition to the costs described above, sexual individuals must also cope with the cost of mating itself, which is in fact a whole set of costs that are difficult to quantify. Sexually reproducing organisms invest much time and energy in securing mates and orchestrating the activity of two individuals which must cooperatively produce young. Much of the ornamentation and behaviour we observe in animals serves these functions. In general, the social interactions required for sexual reproduction cost females time, energy, a heightened probability of catching communicable disease, occasional injury by suitors and even increased predation risk as a result of the conspicuousness of attending males. For males, combat and display behaviour, as well as secondary sexual characteristics maintained by sexual selection, may constitute an additional cost.

It is not possible to sum these costs and give them a precise value, but they are clearly substantial. So, how could sex ever evolve and become predominant in a population when it incurs such heavy costs in competition with asexual reproduction, which incurs none of these costs? The only possibility is for sexual reproduction to carry considerable benefits. These benefits must not only exist, but they must be large, sometimes twofold or more, for sexual reproduction to become established and resist competition from an asexual offshoot. In the next section, we examine some of the possible benefits of sex.

### 4.4.2 THE BENEFITS OF SEXUAL REPRODUCTION

Although reshuffling the genome through recombination may break up valuable combinations of genes, it can also have beneficial effects.

❍ How could recombination benefit an organism's evolutionary fitness?

● Recombination produces offspring, each of which has a unique combination of its father's and its mother's genetic material. In other words, it creates variability.

The question that must now be asked is: what is so valuable about variability? After all, the maternal genotype must have done rather well to have survived to maturity and to have reproduced. Why break up a successful gene combination and incur all the costs of sex in order to gamble on novel genotypes? The answers to these questions have not yet been resolved, but several hypotheses have been suggested, which fall into two categories:

- Sexual reproduction persists because it confers long-term benefits on *populations* or *species*.
- Sexual reproduction is maintained because it confers short-term advantages on *individuals*.

We shall look at each of these ideas in turn.

#### THE LONG-TERM ADVANTAGES OF SEXUAL REPRODUCTION

R. A. Fisher was the first to suggest that the advantage of sexual reproduction lies in the capacity for sexual populations to make a rapid evolutionary response to environmental change. Figure 4.17 illustrates how this process can happen. Suppose that a sexual and an asexual population both have genes $A$ and $B$ at two loci. Two mutations, $A'$ and $B'$, arise in different individuals and are advantageous in the environment where the two populations live. The asexual population eventually consists of $A'B$ and $AB'$ individuals, because in this population, there would be no way for an $A'B'$ individual to arise unless the B gene mutated to $B'$ in a descendant of an $A'B$ individual or the A gene mutated to $A'$ in a descendant of an $AB'$ individual. If mutations are rare, production of an $A'B'$ individual may take a long time. However, in a sexually reproducing population, after $A'$ and $B'$ have arisen in different individuals, recombination can bring them together in the offspring of a mating between these individuals. Recombination would therefore allow favourable mutations to appear together in the same individual sooner than they would if they had to occur sequentially in a single asexual lineage. If they conferred an added advantage when they occurred together, then they would spread through the sexual population more rapidly than through the asexual one. From this reasoning, it follows that a sexual species can evolve faster than an asexual one.

Although Fisher's argument seems plausible, the rate of mutation must be fast in order for it to work. If favourable mutations occur rarely, each becomes fixed in a population (all individuals carry it) before the next one arises and so both sexual and asexual populations evolve at the same rate. New mutations always arise in individuals that already carry the previous mutation (because the previous mutation is fixed in the population and therefore appears in every member of the population). In the above example, the $B'$ mutation arises in an $A'B$ individual in both sexual and asexual populations. However, if favourable mutations arise frequently then Fisher's theory does work.

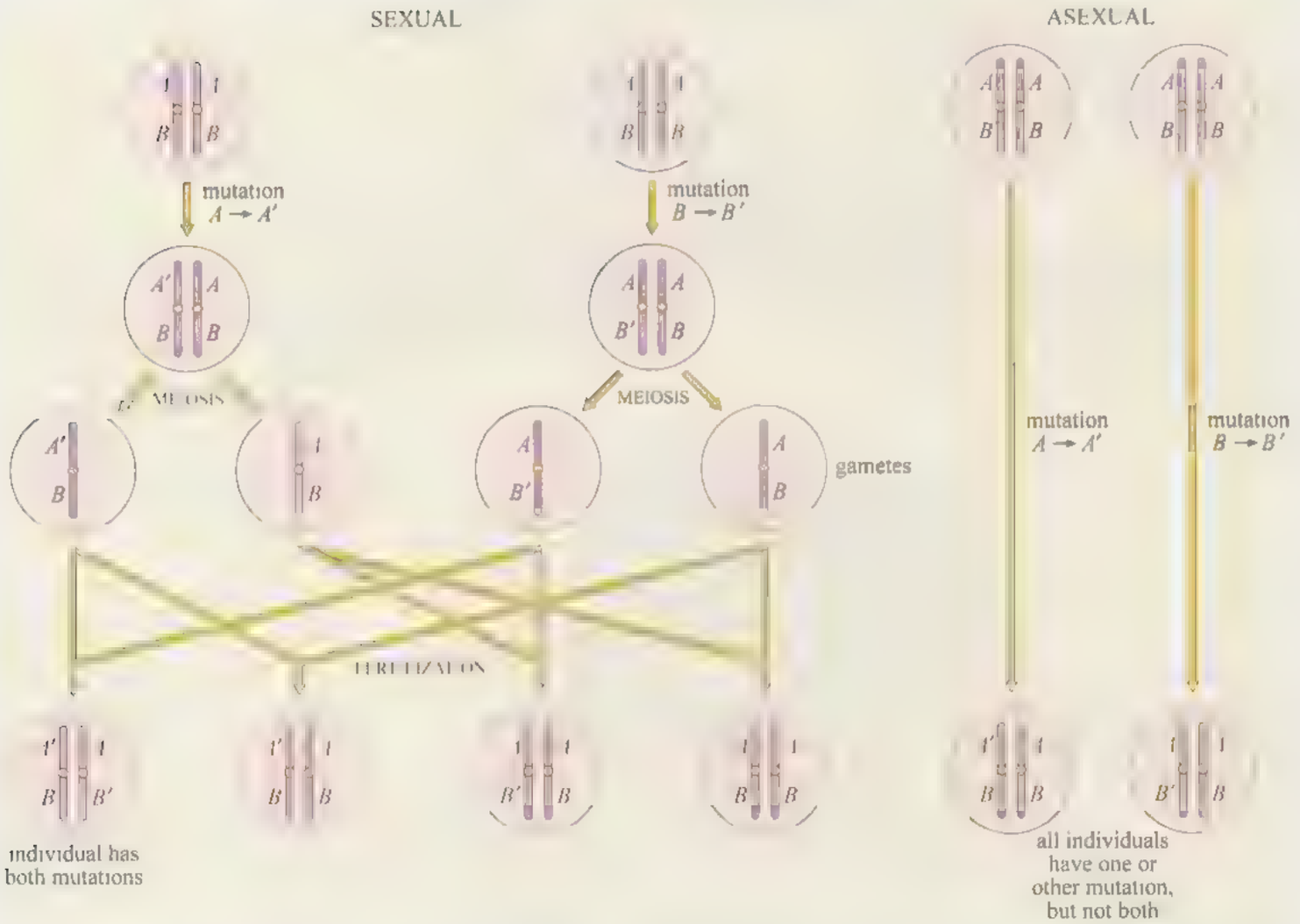



**Figure 4.17** Advantages of sexual reproduction: sex accelerates the rate of evolution because favourable mutations can be brought together in the same individual much faster than in an asexual population.

Muller also pointed out that sexual populations are not only able to evolve faster but are better able to *resist* evolutionary change too. Most mutations damage fitness, and when such a mutation occurs in an asexual line, all descendent individuals are stuck with it; it is not possible for an individual free of the deleterious mutation to appear. Muller evoked the image of a ratchet wheel, clicking inexorably forward with each deleterious mutation accumulated, with no way ever to eliminate errors. At any given time, Muller envisaged a population

that included individuals that carry no mutations, individuals that carry one mutation, individuals that carry two mutations, and so on (Figure 4.18a). Because the population is asexual, Muller thought of the groups as distinct sub-populations. The number of individuals in each group may be small, depending on the size of the whole population. The members of the group with zero mutations on average would enjoy the highest fitness; but if this group is small, then chance events may prevent all individuals reproducing. If this happens, the zero-mutation group dies out, and individuals of the one-mutation group now have the highest fitness (Figure 4.18b). The only way the zero-mutation group can reappear is if an individual in the one-mutation group sustains a mutation that returns the mutant allele to its original form (a back-mutation), which is an unlikely event. The one-mutation group, if it is small, may also be lost by chance in a single generation, leaving the two-mutation group as the highest-fitness individuals (Figure 4.18c). The loss of a small group by chance can occur through genetic drift (Section 3.3.2) and the loss of a group through drift is much easier than its recreation by back-mutation. As the 'ratchet' clicks away, the highest-fitness groups are lost, one by one, from the population, so that over time, the fitness of the population declines. The burden imposed by the accumulation of mutations is known as **genetic load**. When the genetic load becomes too high, the population becomes extinct.

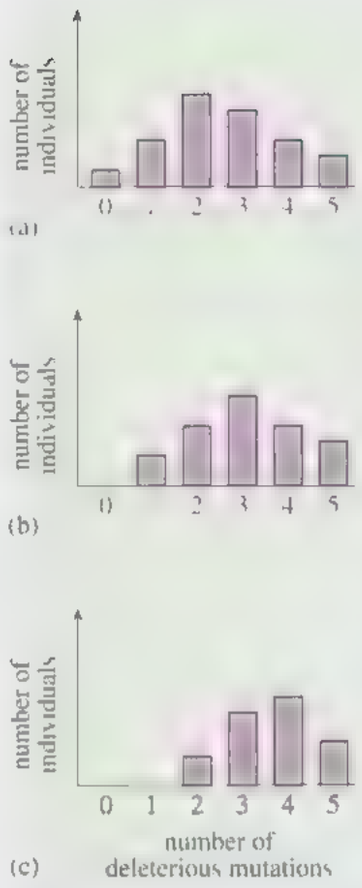



**Figure 4.18** Muller's ratchet. Asexual populations accumulate deleterious mutations. Each of the histograms (a)–(c) shows a snapshot of a hypothetical finite asexual population. In any given generation, the class with the fewest deleterious mutations may be lost (by genetic drift).

Sex breaks the ratchet, because recombination allows sexual strains to 'edit' the genotype. In a sexual population, two individuals with deleterious mutations may produce, by recombination, an individual with no deleterious mutations.

Further research showed that the most critical parameter of Muller's ratchet was population size. In populations of less than ten individuals, genetic drift is a potent mechanism of evolution and the ratchet turns rapidly. In populations of more than 1000, drift is weak and the ratchet does not turn. Some evidence that Muller's ratchet actually operates in organisms is provided by studies on invertebrates that normally alternate between sexual and asexual generations, e.g. aphids. If the experimenter intervenes to prevent sex, repeated asexual reproduction may be accompanied by an accelerating decline in fitness in the asexual lines, apparently due to the accumulation of deleterious mutations.

❍ Why is the view that sex is of value because it accelerates the rate of evolution of species or populations open to criticism?

● Because it provides no explanation of how sexual reproduction can become *established* in a population in the first place.

Because of the twofold advantage to females of reproducing parthenogenetically, whenever a parthenogen arises in a sexual population, it can replace the sexual population that gave rise to it simply by the rapid spread of parthenogenetically transmitted genes. Only in the long term do sexual forms supplant their asexual competitors because they evolve more rapidly. However, any population that reproduces sexually must resist invasion by parthenogens in the short term, and it can do so only if there is some short-term *individual advantage* to sex.

### SEX AND INDIVIDUAL ADVANTAGE

Evidence that there must be some short-term advantage to sex and that casts some doubt on the importance of rapid evolution of populations in the maintenance of sex comes from an argument put forward by Williams in 1975, which is known as 'the balance argument'. Many plants, aphids, sponges, rotifers and water fleas (*Daphnia*) can reproduce both sexually and asexually, depending on conditions. Many of these species time their sexual reproduction to periods of environmental uncertainty, and reproduce asexually when conditions are more stable, i.e. they are cyclical parthenogens. Williams' point is that when an individual such as an aphid reproduces sexually, it must be advantageous to that individual, otherwise the aphid would have reproduced asexually (since it has a 'choice'). Both sexual and asexual reproduction must have 'balanced' advantages to maintain them in the species life cycle, otherwise the inferior method would be lost. The persistence of occasional sex in these species therefore suggests that the costs of sex are met with benefits in individual fitness.

Since the balance argument was first raised by Williams, there have been several mechanisms proposed for the maintenance of sex in populations in the short term. We will consider two of these, sib-competition models, suggested by Williams, Maynard Smith and Bell, and the Red Queen hypothesis, which was put forward by Seger and Hamilton. Both hypotheses focus on the advantages associated with producing genetically variable offspring.

### SIB-COMPETITION MODELS

The essential feature of sib-competition models is that individuals produce a large number of offspring which compete with one another for limited resources. One hypothesis that relies on competition between siblings is Bell's 'tangled bank model', which is based on a reference to a diverse, multi-species environment, described in Darwin's *The Origin of Species*. Darwin realized that if conditions were the same everywhere, a few species would come to dominate any habitat. Yet, when he surveyed an area of about a square metre of his own lawn, he counted over 20 different species of plant. Darwin's description of a tangled bank emphasized the fact that, within a small area, many different microhabitats exist, which differ with respect to factors such as shade, moisture, exposure to wind, suitable germination or nesting sites, availability of mates, etc. So, within each patch of habitat, there is a spectrum of resources that can be partitioned among individuals in that patch.

If the habitat is underpopulated, it may not matter just how an organism is specialized to deal with the environment. However, if the habitat is crowded, small differences among individuals in resource utilization may begin to matter. Organisms that reproduce asexually may be at a disadvantage in a crowded, heterogeneous habitat because they are superior in only one microhabitat. They produce multiple clones of themselves, each of which competes for the same resources. Under these conditions, siblings compete among themselves and the more individuals there are, the more intense this competition.

In contrast, sexual species produce offspring with different genotypes which are able to exploit small-scale differences in the environment. Because their genes differ, sexual organisms can utilize the resources available to them without

competing directly with one another. This phenomenon is known as **frequency-dependent selection** because the advantage of a trait depends on how many other individuals have it. In crowded habitats, where competition is important, frequency-dependent selection leads to a situation where there is a balance between the numbers of organisms with alternative traits: if individuals with a certain rare trait have an initial advantage, the number of offspring carrying that trait increases. But as soon as they begin to compete with one another, the value of the trait drops until there is no advantage. In other words, rarer types do better than common ones (because competition is less intense) but their advantage disappears as the density of once-rare types increases.

A second model that relies on sib competition is known as the **lottery model**. Here again, the environment is considered as a series of patches, but there is heterogeneity *between* patches rather than within them. Each patch is colonized by progeny from several parents but only one individual — the one best adapted to that particular patch — survives to reproduce. Since the offspring of sexual parents have varied genotypes, the progeny from a single individual can survive on several different types of patch. Offspring from asexual parents, on the other hand, can only survive on patches similar to the parental patch. Under this hypothesis, each patch can be considered as having a single winning lottery number. A sexual parent has many different lottery tickets with different numbers (different offspring genotypes). An asexual parent, on the other hand, holds numerous tickets with the same number. The probability is therefore greater that the sexual female has the ticket with the winning number for each patch.

The sib-competition and lottery models are concerned with the advantage of sex in dealing with spatial heterogeneity and focus on the unpredictability of the physical environment. In contrast, the second hypothesis we will discuss focuses on variation in an organism's biotic environment and considers how individuals deal with heterogeneity over time.

### THE RED QUEEN HYPOTHESIS

The **Red Queen hypothesis** for the maintenance of sex states that sexual reproduction enables individuals to escape from parasites. The coevolution between parasites and hosts is thought to generate environmental change at a speed that renders sex advantageous in the short term. The 'environment' for the parasite is the host's resistance mechanism; for the host, the 'environment' consists of the parasite's method of penetrating its defences.

The Red Queen hypothesis was originally formulated by Leigh Van Valen in 1973 to explain the fact that, over evolutionary time, species are continuously becoming extinct and replaced by other species. The hypothesis took its inspiration from the scene in *Through the Looking Glass* by Lewis Carroll in which Alice and the Red Queen are running as quickly as they can and yet make no progress: however fast they move, they cannot outrun their surroundings. So, for proponents of the Red Queen hypothesis, life is an evolutionary race: you must sprint as fast as you can just to keep up, never mind getting ahead. The hypothesis in effect states that because the relative importance of factors such as predation, competition and parasitism varies over time, species have to keep changing in order to 'stay in the same place'.

DNA molecule (Section 5.7.3), the diverging replication forks meet up on the other side of the circular DNA, yielding two separate copies of the double-stranded DNA (each containing one template strand and one newly synthesised strand). In eukaryotes, the replication forks originating from multiple origins meet up with each other, ultimately resulting in two separate copies of each linear chromosomal DNA molecule.

The animation at the end of Activity 5.1 (Part 1) showed a dynamic view of the replication complex in action, based on structural studies of the proteins in the complex. In this case, the parental DNA helix can be seen entering the replication complex from the left (this image is in the opposite orientation to Figure 5.4a), the leading strand is exiting downwards, and you can see the template for the lagging strand being spooled out in a big loop as Okazaki fragments are synthesised (Figure 5.5).

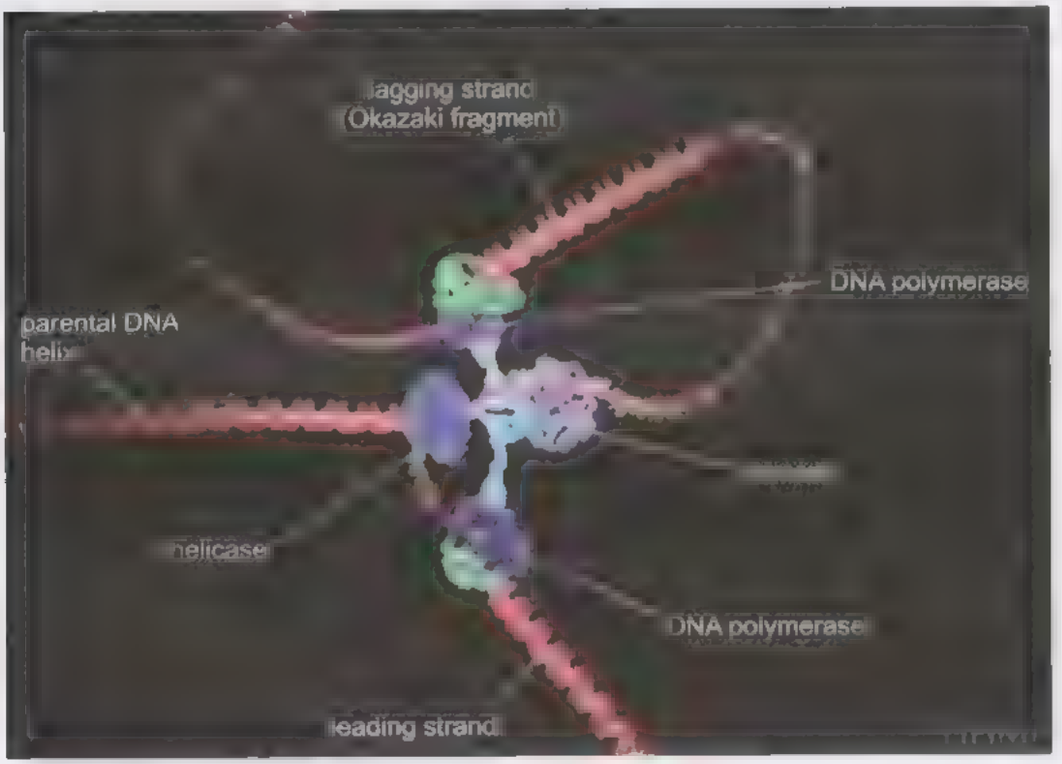



**Figure 5.5** A model for the DNA replication complex, showing how the DNA strands form a complicated looping structure around the site of synthesis.

The ability of isolated DNA polymerases to carry out DNA synthesis on a purified DNA template in a test tube is widely exploited in molecular biology, most notably in the polymerase chain reaction, one of the most influential molecular biology techniques developed in recent decades (Box 5.1), and in DNA sequencing (Box 5.3).

## Box 5.1 The polymerase chain reaction

The **polymerase chain reaction (PCR)** is a technique that permits the rapid amplification of a DNA fragment from a vanishingly small amount of starting material isolated from cells. It yields sufficient quantities

(micrograms) of accurate DNA copies for techniques like DNA sequencing and gene cloning (described later in this chapter) and is a particularly powerful and versatile technique – indeed, its main inventor Kary Mullis earned a Nobel Prize.

PCR makes use of the naturally occurring enzyme activities that replicate the double-stranded DNA in cells. To synthesise DNA *in vitro*, a number of biomolecules are required:

- a double-stranded template DNA
- DNA polymerase (the enzyme that catalyses the addition of nucleotides to a new DNA strand)
- the four dNTPs (dATP, dGTP, dCTP and dTTP)
- two oligonucleotides (short lengths of single-stranded DNA, generally around 20 nucleotides long) to act as primers – one for each DNA template strand.

■ Why does DNA synthesis require oligonucleotides to act as primers?

Recall that the DNA polymerase cannot initiate the synthesis of a DNA strand. Instead, polymerisation begins by addition of nucleotides to an existing short nucleic acid primer annealed to the template strand.

There are two key aspects of PCR that make it so useful. Firstly, the sequences of the two oligonucleotide primers determine exactly which section of a DNA molecule will be copied by the polymerase. The specificity of complementary base pairing between a primer and its target DNA means that a particular sequence may be specifically targeted and amplified from a much larger mass of other DNA sequences. Secondly, repeated rounds of DNA synthesis, in which each newly synthesised section of DNA itself becomes a template, lead to an exponential increase in the quantity of the specific DNA product.

Each round of DNA synthesis has three phases:

1 *denaturation*: separation of the strands of DNA to yield two single-stranded DNA templates
2 *primer annealing*: base pairing of the oligonucleotide primers to their complementary sequences
3 *elongation*: synthesis of two new DNA strands by DNA polymerase, initiated from the oligonucleotide primers.

The method is remarkably simple. However, it is first necessary to know part of the DNA sequence of the region to be amplified in order to chemically prepare the short synthetic DNA primers for initiating the reaction. The two primers are complementary to the ends of the region that is to be amplified, and are usually about 20 nucleotides long. The DNA polymerase used for PCR is a thermostable (heat-resistant) enzyme, derived from a bacterium adapted to living in high temperature environments: for example, the commonly used *Taq* DNA polymerase is

derived from the bacterium *Thermus aquaticus* which inhabits hot springs.

A tube containing the reaction mixture of target DNA, primers, nucleotides and *Taq* polymerase is first placed at 94 °C, which denatures the template DNA (disrupting the weak hydrogen bonds between base pairs) and separates it into the two single strands. The reaction is then transferred to a lower temperature (usually around 55–58 °C; the exact temperature depends on the sequence of the primers) to allow the primers to anneal (i.e. bind) to the template DNA strands via complementary base pairing. Finally, the reaction is incubated at 72 °C for polymerisation (elongation) to occur (*Taq* polymerase functions at an optimal temperature of 72 °C). This represents one cycle of denaturation, primer annealing and elongation (Figure 5.6a). The thermostable *Taq* polymerase is resistant to denaturation at 94 °C (a temperature that irreversibly damages most enzymes), so the reaction mix can be subjected to 30 or 40 of these cycles without the addition of fresh enzyme. Repeating the cycle 30–40 times causes amplification of the short target sequence (Figure 5.6b). As long as the dNTPs and primers are present to excess, the amount of target amplified grows exponentially. By the later cycles, the primers and free dNTPs begin to be depleted as they are consumed by synthesis, and the amplification rate slows.

## Summary of Sections 5.1 to 5.3

- All organisms use DNA as their genetic material. The double-stranded structure of the DNA double helix, with its complementary base pairing between the bases of the two strands, is key to the stability of DNA, the mechanism for efficient and accurate DNA replication (copying) and its ability to encode genetic information.
- DNA replication involves a large protein complex (the replication complex) comprising a number of proteins including: DNA helicase, which unwinds and separates the two DNA strands; primase, which provides RNA primers that anneal to the DNA template strands and initiate new-strand synthesis; and DNA polymerase, which links nucleotides to the new DNA strands.
- DNA synthesis is unidirectional, 5′ to 3′ (so the template is always 'read' 3′ to 5′), which has implications for replication. While the leading strand can be synthesised as one continuous DNA molecule, the lagging strand is synthesised in short sections, which are finally joined into a complete strand through the action of DNA ligase.
- DNA polymerases have a proof-reading function which monitors the newly synthesised DNA strand for misincorporation of nucleotides.
- Thermostable forms of DNA polymerase can be used *in vitro* to carry out the polymerase chain reaction (PCR), a method of amplifying a specific section of DNA using synthetic primers to initiate DNA synthesis.

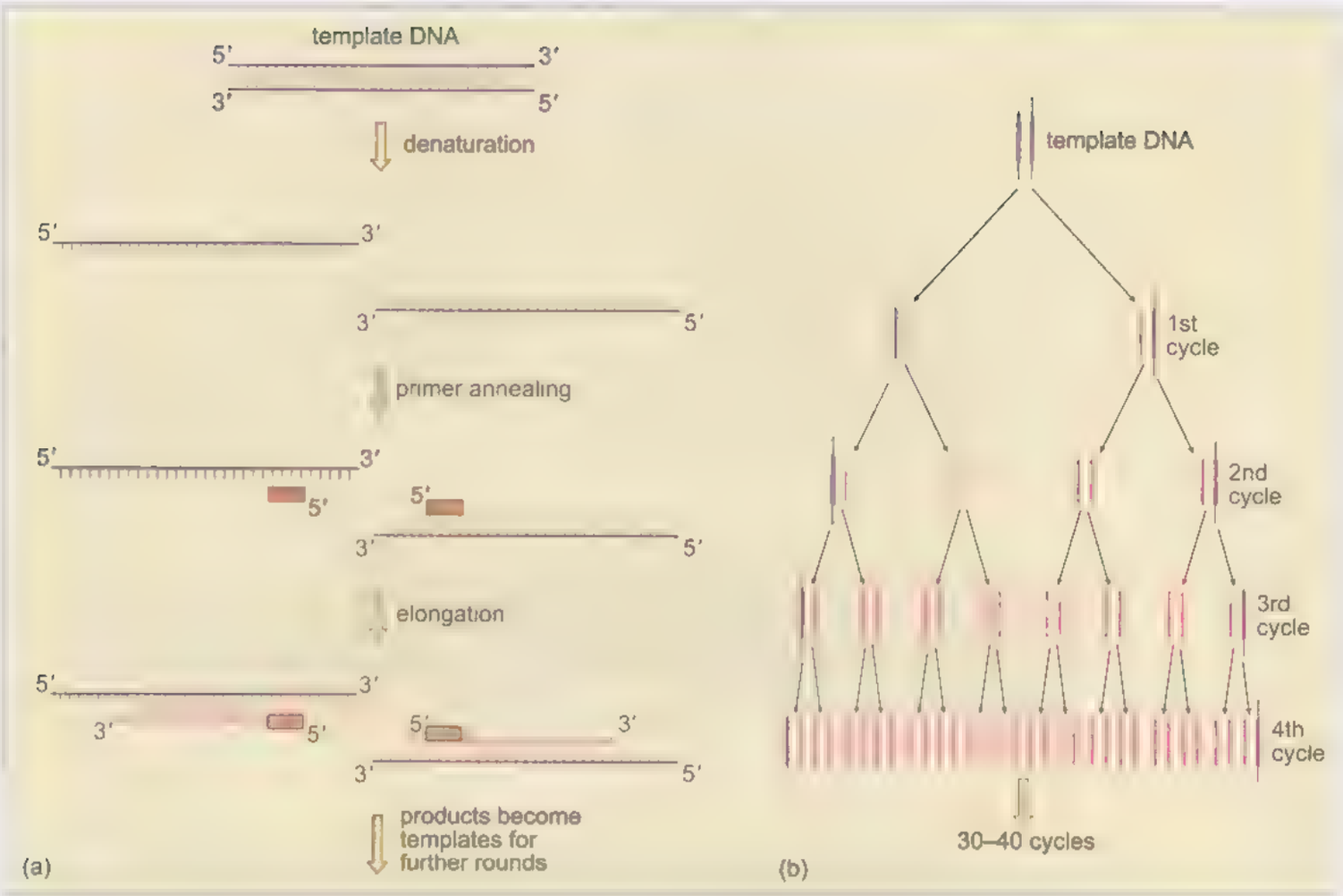



**Figure 5.6** The polymerase chain reaction. (a) Each cycle consists of three phases: denaturation, primer annealing and elongation. During denaturation, the DNA is heated to approximately 94 °C to separate the strands. In the annealing phase, the temperature is reduced, permitting the short DNA primers to anneal to complementary sequences in the template strand. During elongation, the temperature is elevated to 72 °C, the optimum temperature for thermostable DNA polymerase, and DNA synthesis occurs. (b) After the first round, the product itself becomes a template so that after several cycles of PCR, there are many copies of a short double-stranded section of DNA. Typically, 30–40 cycles of amplification will be carried out, leading to an exponential increase in the mass of the target sequence.

## 5.4 DNA repair

Any unrepaired errors that occur during DNA replication or as a result of damage to DNA will give rise to mutations, permanent changes in the genomic DNA sequence of the next generation of cells. If an error is not repaired, it will be copied when the cell divides and the DNA of subsequent cells will contain the change. Clearly, when this happens, the information carried by the DNA is also changed and in some cases this may have profound effects on expression of gene products (Section 5.6). However, all organisms have a number of DNA repair mechanisms.

## 5.4.1 Correction of errors in replication

While the innate proof-reading capacity of DNA polymerase ensures a very high degree of accuracy during DNA replication, occasionally incorrect nucleotide incorporation is overlooked. Cells are able to correct nucleotide mismatches that fail to conform to normal base pairing rules by a process known as **mismatch repair (MMR)**.

Mismatch repair enzymes can recognise the newly synthesised strand and ensure that, where there are mismatches, the misincorporated nucleotide in the new strand (not the template strand) is the one replaced. In some bacteria, DNA becomes methylated some time after replication (methyl groups are added to some of the nucleotides) and the MMR enzymes appear to recognise which is the newly synthesised DNA strand by its lack of methylation just after replication. The recognition mechanism in eukaryotes is less clear, but may involve detection of temporary single-stranded breaks, known as 'nicks', in the newly synthesised DNA strand. Recall that A T and G C base pairs are of equal length and allow the sugar phosphate backbones of the DNA duplex to be maintained a certain distance apart.

- How might repair enzymes detect a base pairing mismatch, such as A:C?

  The mismatched base pair is an inappropriate length and will cause a distortion in the DNA helix.

A complex of MMR enzymes nicks the new strand close to the distortion, unwinds a short stretch of the helix, then an *exonuclease* (an enzyme that removes nucleotides one at a time from the end of a polynucleotide chain), starting at the nick, degrades a section of the new strand, including the mismatch. A DNA polymerase then fills in the gap left in the strand by synthesising new DNA in the 5′ to 3′ direction, using the other, intact DNA strand as template. The enzyme DNA ligase completes the repair by joining the sugar phosphate backbone of the newly synthesised section of DNA to the end of the strand, so that there are no remaining gaps (you encountered DNA ligation earlier in this chapter, in connection with joining of the short, single-stranded Okazaki fragments during lagging-strand DNA synthesis).

## 5.4.2 Repair of damaged DNA

Although most mutations are believed to be caused by replication errors, organisms are also subject to continual assault by a variety of agents that can damage biological macromolecules, including DNA. These agents include electromagnetic radiation such as ultraviolet light, X-rays and gamma-rays; toxic chemicals in the environment, such as those found in cigarette smoke, and also chemicals generated as a by-product of normal biochemical activity, such as reactive oxygen species. Each of these agents produces characteristic types of molecular damage to DNA (often referred to as molecular *lesions*) that can affect function. These lesions are corrected by a number of different repair processes, many of which are able to accurately recognise and repair errors in one strand using the intact complementary DNA strand as a template. Several DNA repair processes exist in both prokaryotic and eukaryotic cells,

Reactive oxygen species (ROS) are highly chemically reactive molecules containing oxygen atoms with single unpaired electrons. ROS are generated by natural metabolic processes within cells, and are constantly removed to prevent them from damaging other molecules.

and many of the proteins involved have been highly conserved in many different species throughout evolution.

### *Base excision repair*

As its name suggests, **base excision repair (BER)** is a cellular mechanism whereby single damaged bases are removed from the DNA. Damage to bases is often a spontaneous event caused by normal chemical reactions in the cell. A common example is the *deamination* of cytosine: the replacement of an amino group ($-NH_2$) in a cytosine base by a carbonyl group (C=O), which converts it into another type of base called uracil (Figure 5.7a).

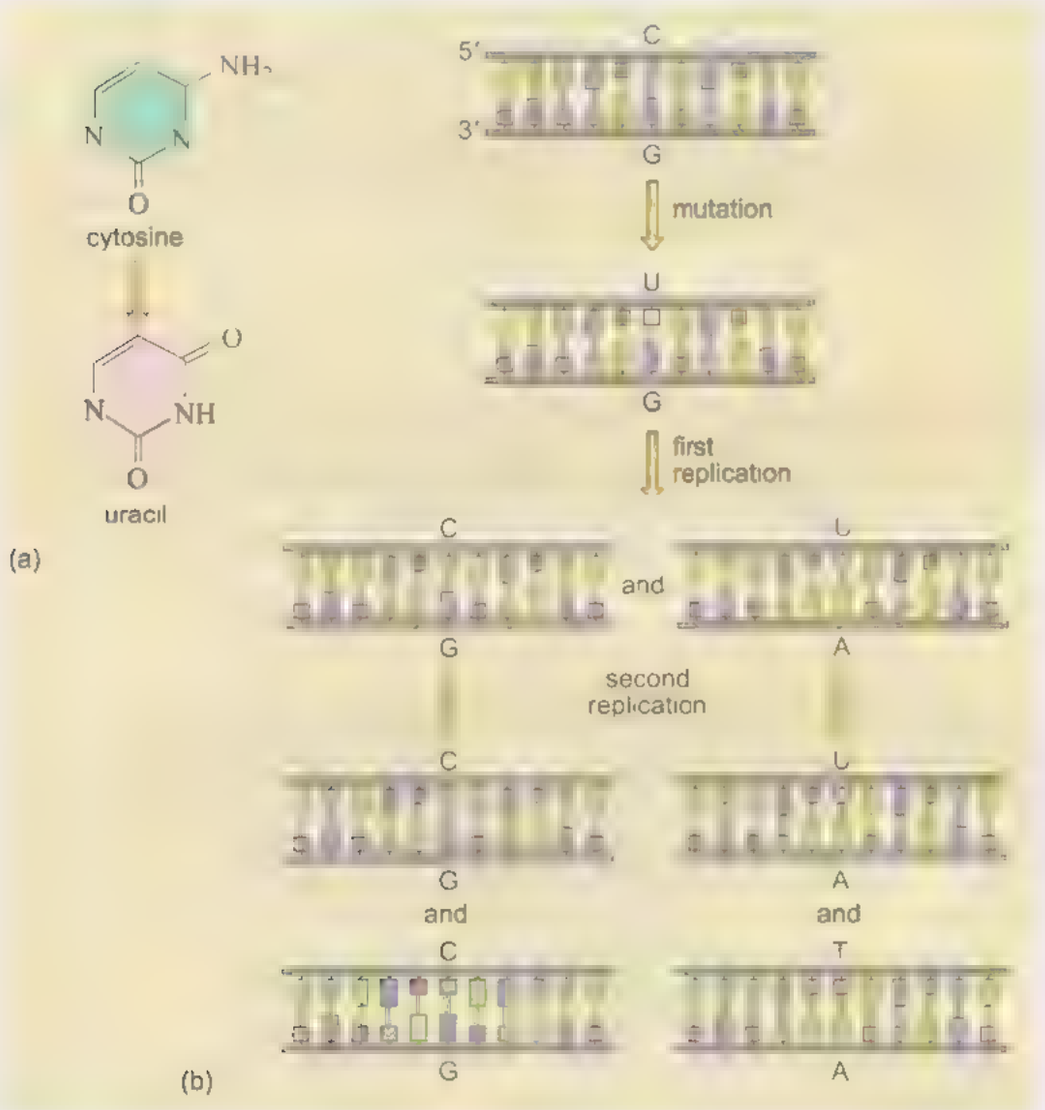



**Figure 5.7** (a) Deamination of cytosine forms uracil, which can base-pair with adenine, but is not normally found in DNA. (b) If the uracil is not removed prior to DNA replication, it can lead to a replacement of a C:G base pair with a T:A base pair.

Uracil is normally found in RNA (where it replaces thymine) but not in DNA, and its presence in DNA is problematic for DNA replication, since uracil pairs with adenine.

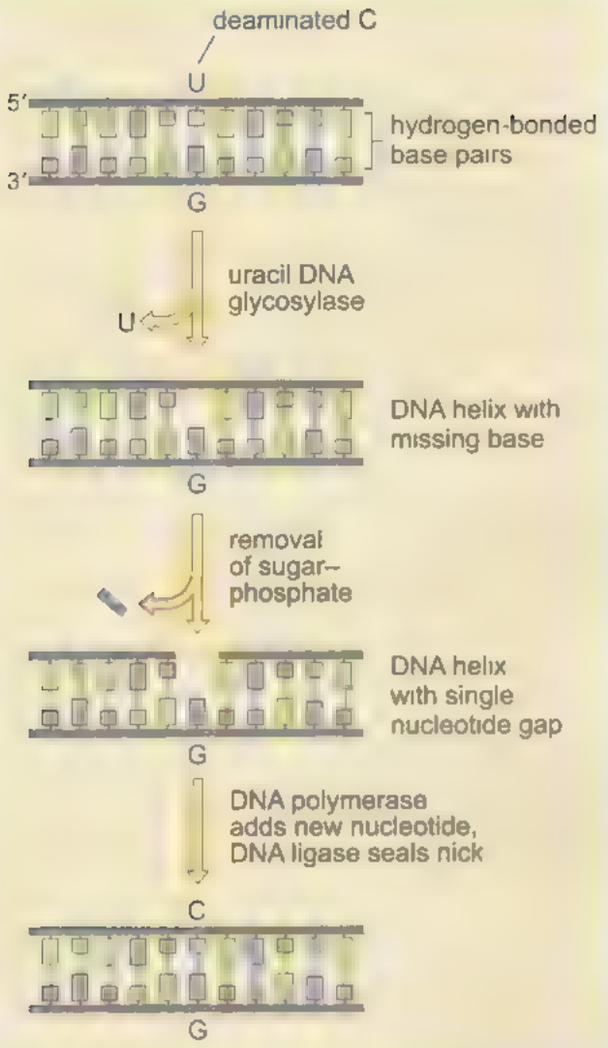



**Figure 5.8** The base excision repair (BER) pathway which removes damaged bases. In this example, deamination has converted a cytosine into a uracil residue. The uracil is removed, leaving an abasic site, which is itself cleaved from the DNA, leaving a short gap in one strand. The gap is repaired by DNA polymerase and DNA ligase.

- What might be the biological consequences of conversion of C into U by deamination?

  C normally base-pairs with G, but U base-pairs with A, so replication using DNA template strands containing U in place of C will lead to misincorporation of A in place of G in the new DNA strand (Figure 5.7b).

Two other bases, adenine and guanine, can also be affected by deamination. Base excision repair is initiated by the action of DNA glycosylases, enzymes that catalyse the cleavage of a base from the sugar -phosphate backbone. There are several different DNA glycosylases with specificity for particular damaged bases. In the example shown in Figure 5.8, deamination of cytosine has resulted in the formation of a uracil residue. The incorrect base is recognised by uracil DNA glycosylase, and cleaved from the sugar–phosphate backbone leaving a deoxyribose sugar lacking a base, known as an *abasic* site. In a second step, the deoxyribose sugar–phosphate is removed by nucleases, and the gap is repaired by the combined action of DNA polymerase, which adds the correct nucleotide, and DNA ligase, which rejoins the sugar–phosphate backbone, leaving no gap.

Another common type of chemical damage is *depurination*, the complete loss of a guanine or adenine base from the sugar–phosphate backbone, leaving an abasic site. This type of damage can also be corrected by the BER pathway (Figure 5.8).

### *UV damage and nucleotide excision repair*

The Sun's rays include invisible ultraviolet (UV) light, which is categorised into UV-A, UV-B and UV-C on the basis of wavelength. All UV exposure causes damage, but UV-B (wavelength 315 nm–280 nm) is particularly damaging to DNA. Exposure to UV-B light leads to the formation of pyrimidine dimers, where adjacent cytosine or thymine residues undergo photochemical reactions that covalently link the two bases (Figure 5.9a).

Thymine dimers interfere with normal base pairing, and also induce a distortion, or 'kink', in the DNA double helix (Figure 5.9b). These changes in the DNA structure prevent accurate DNA replication and transcription. Thymine dimers and similar lesions are repaired by a process known as nucleotide excision repair (NER). NER is a complex process, which in humans relies on the products of at least 30 genes, operating in as many as 18 different protein complexes.

The module learning resources include an interactive model of a thymine dimer, which allows you to view this distortion in three dimensions.

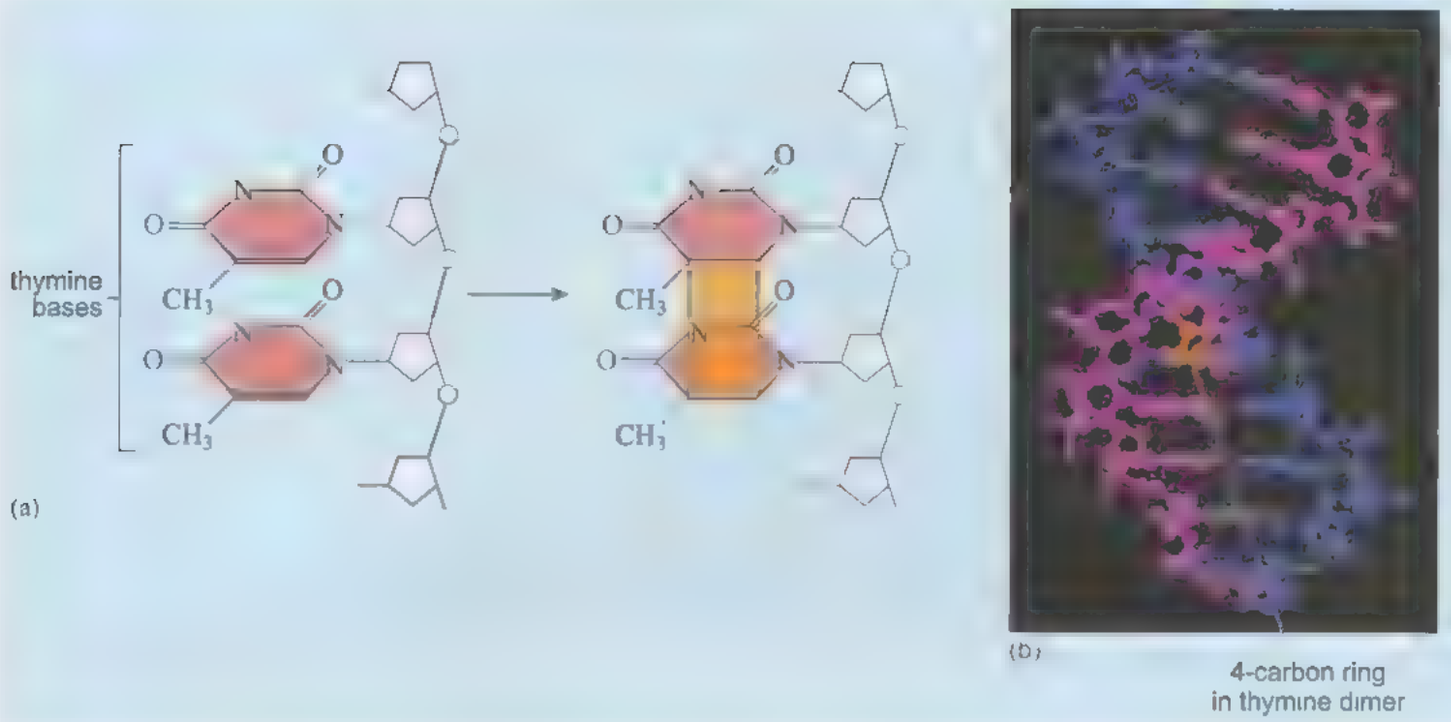



**Figure 5.9** (a) A thymine dimer is formed by the covalent linkage of a thymine residue with a neighbouring thymine, forming a 4-carbon ring. (b) The thymine dimer forms a kink in the double helix, clearly seen by looking at the alignment of base pairs above and below the thymine dimer in this molecular model (the 4-carbon ring is highlighted in orange).

In outline, repair of thymine dimers by NER is initiated when the distortion or 'kink' in the double helix caused by the presence of the thymine dimer is recognised by a complex of repair proteins. The DNA strand containing the dimer is cleaved on either side of the lesion by nucleases, and the short single-stranded section containing the lesion is released by the unwinding and separation of the two strands by DNA helicase (Figure 5.10). This leaves a short gap in the strand that is filled by synthesis of new DNA using the intact DNA strand as a template. The gap remaining at the end of the new single-stranded section is finally sealed by DNA ligase.

- What might be the consequence of UV-B exposure for the cells of an individual with an inherited defect in the NER repair pathway?

  Individuals incapable of repairing UV-induced damage such as thymine dimers would accumulate DNA damage and mutations in the cells in the areas of their body that are exposed to sunlight.

Xeroderma pigmentosum (XP) is a human genetic disorder in which affected individuals are highly sensitive to UV-induced DNA damage due to defects in NER. XP patients typically suffer from extreme sunburn, damaged skin and a high frequency of skin tumours in areas of skin that are exposed to sunlight. The genetics of XP is complex: there are at least seven different types, referred to as XP-A through to XP-G, each caused by mutation of a gene that encodes one of the proteins involved in NER. You will return to a detailed discussion of xeroderma pigmentosum in Book 3.

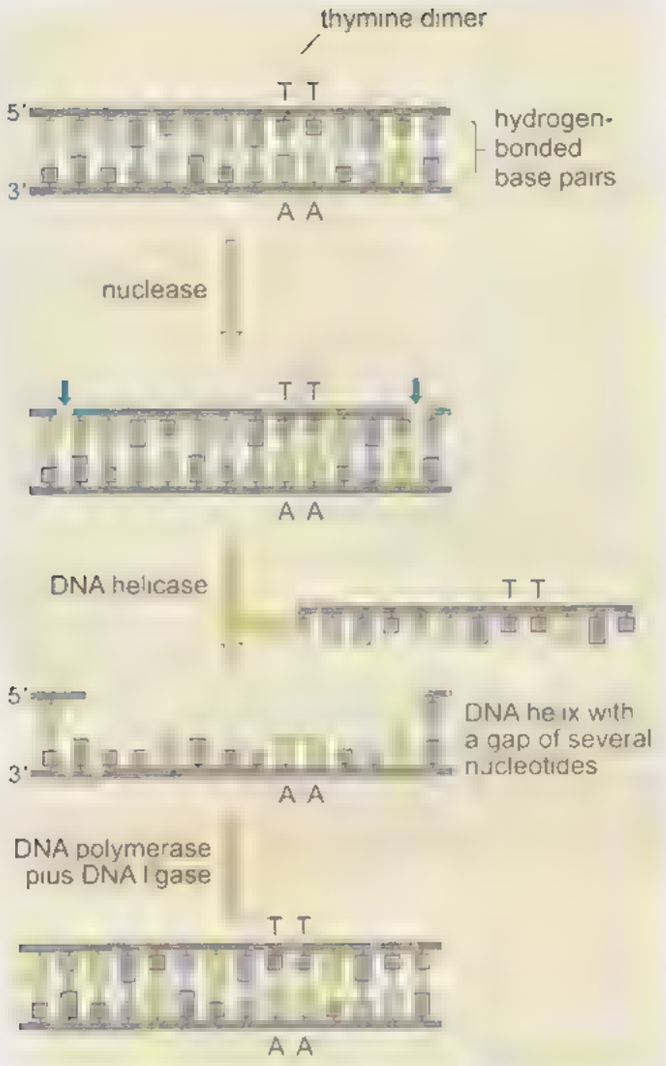



**Figure 5.10** The nucleotide excision repair pathway. The strand containing the lesion, in this case a thymine dimer, is cleaved on either side of the lesion. The cleaved section is detached by the action of DNA helicase, which separates base pairs, and the missing section is replaced by the action of DNA polymerase and DNA ligase.

### *Repair of double-strand breaks*

As described above, MMR, BER and NER all transiently generate short single-stranded sections of DNA which can be efficiently and accurately repaired by DNA synthesis, using the other intact DNA strand as a template for new synthesis. In contrast, the potential consequences of a break in both strands, a double-strand break (DSB), can be rather more severe, and will result in cell death if unrepaired.

- Why might you expect unrepaired double-strand breaks to be severely deleterious to the cell?

  Unrepaired DSBs could lead to fragmentation of the chromosomes.

Double-strand breaks are typically caused by replication errors or highly energetic ionising radiation, including X-rays. They can be repaired by one of

two mechanisms. The first, which is known as non-homologous end-joining, involves a protein complex that simply trims the two damaged ends and ligates them together again (Figure 5.11a). This type of repair is very 'error-prone', because during this process several base pairs of DNA are usually lost before the break is repaired, and if many DSBs have been induced at the same time, there may be incorrect rejoining of DSBs from different places on the chromosome, or even between different chromosomes.

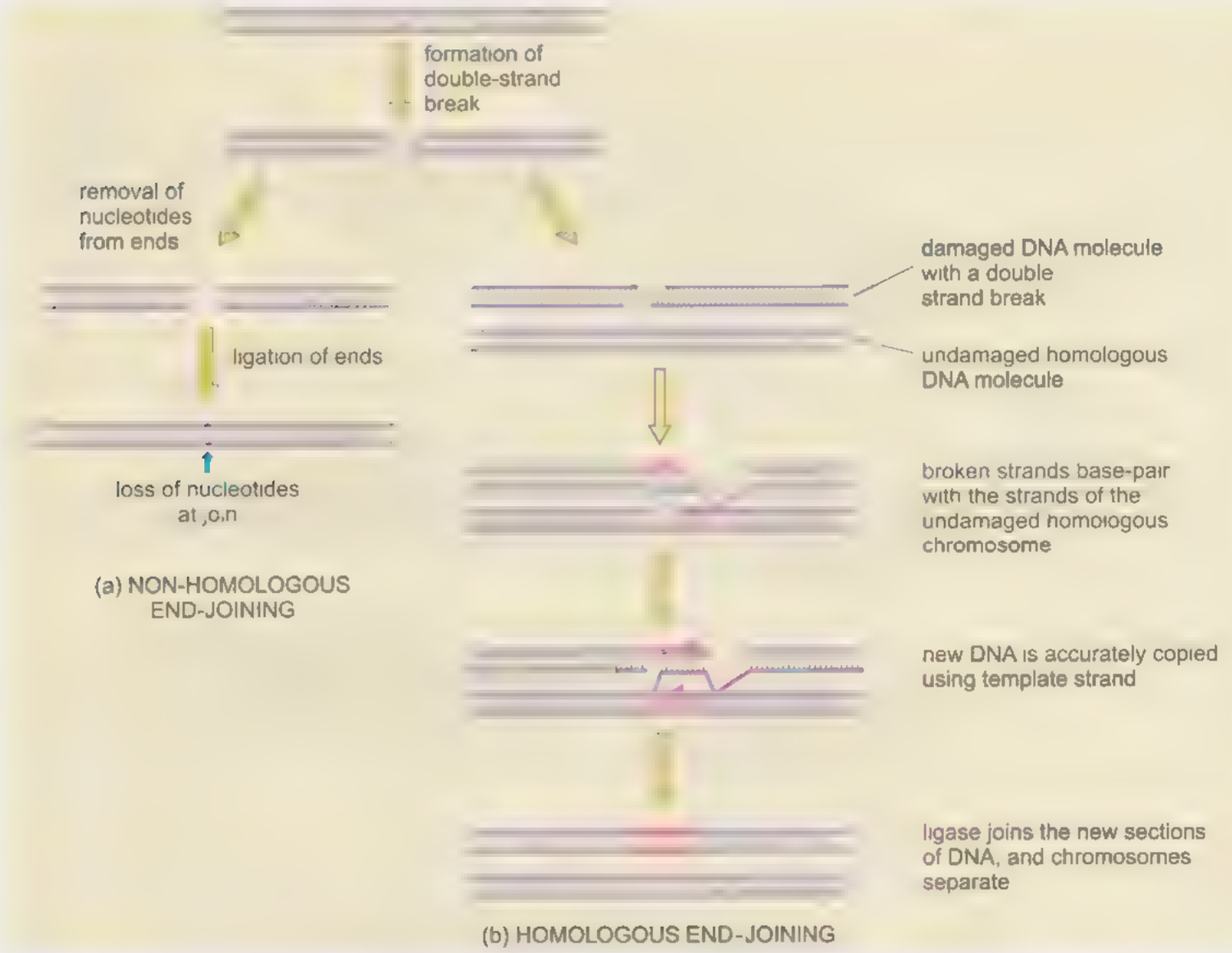



**Figure 5.11** Repair of double-strand breaks. (a) Non-homologous end-joining is error-prone as it does not use a template to ensure the repair is accurate and it also involves the loss of a few nucleotides from the site of breakage. (b) Homologous end-joining is a more accurate mechanism for repairing double-strand breaks. The damaged strands pair with another intact homologous DNA molecule (the homologous chromosome in a eukaryotic cell) which then acts as a template for DNA repair.

In contrast, in a second mechanism, known as homologous end-joining (Figure 5.11b), the damaged DNA molecule pairs with another (undamaged) homologous DNA molecule which acts as a template for DNA synthesis to repair both strands accurately. Recall that the cells of diploid eukaryotes have pairs of homologous chromosomes, so the other (undamaged) chromosome of

the pair can be used as a template for DNA synthesis to repair a DSB (Figure 5.11b). This mechanism is therefore more accurate than non-homologous end-joining, and is generally the cell's preferred mechanism of DSB repair. This repair pathway essentially uses the same molecular mechanism as genetic recombination (crossing over) during meiosis (Section 4.3.1). Indeed, it can sometimes result in a genetic exchange between the two chromosomes.

## 5.5 [illegible] DNA to protein synthesis

Before going on to look at the genome composition of different organisms, this section will first consider the information encoded by genes, and how changes to the nucleotide sequence of a gene can affect its function. In essence, a gene is a sequence of nucleotides that encodes a product – usually a protein. Chapter 6 of this book deals with the process of gene expression in some detail, but for now gene structure will be considered in general terms, and the discussion restricted to protein-coding genes. Activity 5.1 (Part 2) continues the description of the flow of information in the cell, by describing the transcription of the DNA sequence to yield a messenger RNA intermediate, which is then translated to synthesise a polypeptide at the ribosome.

### Activity 5.1 The flow of information in cells (Part 2)

**(LO 5.4) Allow 30 minutes**

The second part of this activity outlines how the information encoded in DNA specifies the amino acid sequence of an encoded polypeptide. Again, you may already be familiar with this process from earlier studies, but are recommended to review this material before proceeding with the chapter.

The sequence of amino acids in every protein in the cell is determined by the sequence of the four DNA bases – adenine (A), guanine (G), cytosine (C) and thymine (T) – in the gene that encodes the protein. The DNA sequence of the gene is first copied by RNA polymerase to produce a single-stranded **messenger RNA (mRNA)** in a process known as **transcription**, beginning and ending at specific start and termination points in the DNA (Activity 5.1, Part 2 and Figure 5.12). In contrast to DNA replication, only one strand of the DNA is used as a template for the synthesis of the mRNA. The DNA template strand is again read in the 3′ to 5′ direction, and the mRNA is synthesised in the 5′ to 3′ direction (Figure 5.12). The structure of RNA is similar to DNA; it has a ribose sugar backbone (whereas DNA has a deoxyribose sugar backbone) and while it also has four bases, these are A, G, C and U, because thymine (T) is replaced by uracil (U) in RNA. Uracil (like thymine) forms base pairs with adenine. RNA, unlike DNA, is usually single-stranded.

The single-stranded mRNA is then used by ribosomes as the template for protein synthesis in the process called **translation** (Figure 5.12 and

Activity 5.1, Part 2). During translation, the mRNA bases are recognised in groups of three, each known as a **codon**, which specify one of the 20 types of amino acid found in proteins.

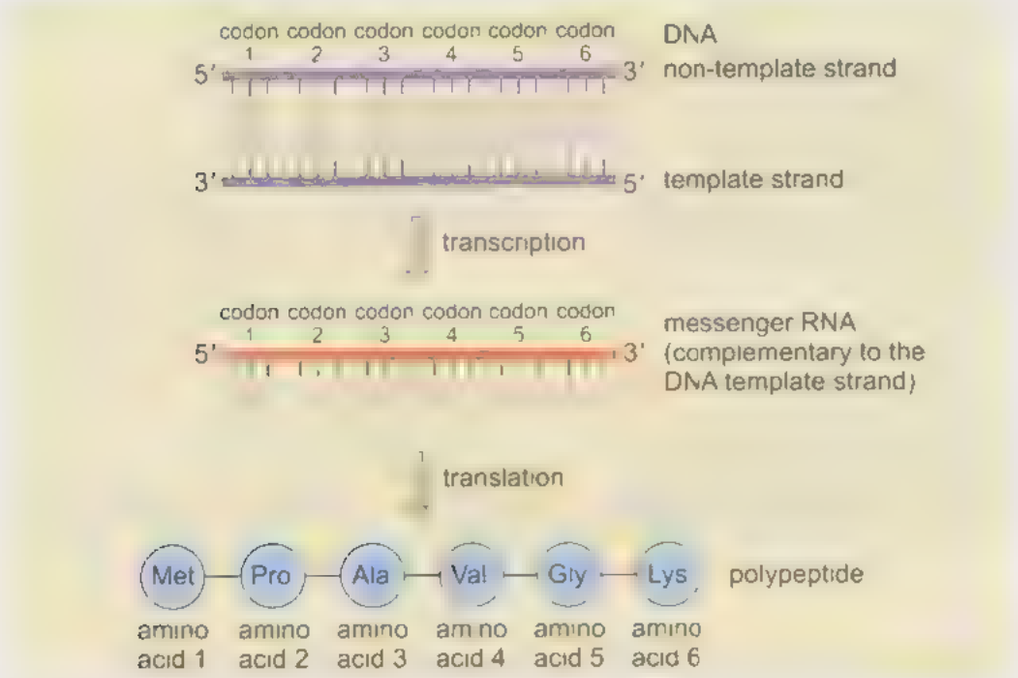



**Figure 5.12** The flow of genetic information. The sequence of the DNA template strand of a protein-coding gene is copied into a single-stranded messenger RNA (transcription). At the ribosome, the mRNA is used as a template for synthesis of the polypeptide chain (translation). Each of the triplet codons in the mRNA message encodes a particular type of amino acid.

The relationship between the triplet base codons in RNA and the amino acids they specify is known as the **genetic code** (Figure 5.13). The four bases in RNA can be arranged in 64 different combinations of three (there are four possibilities for the first base of the codon (U, C, A or G), four for the second, and four for the third; $4 \times 4 \times 4 = 4^3 = 64$). Since there are 64 different codons and only 20 naturally occurring amino acids, it follows that some amino acids have multiple codons; this is sometimes referred to as *redundancy* of the genetic code. There is also a 'start' codon (AUG, which encodes the amino acid methionine) and three 'stop' codons (UAA, UAG and UGA), which signal the termination of translation (Figure 5.13) and which do not encode an amino acid. You will learn about the processes of transcription and translation in more detail in the next chapter.

- Using the genetic code table in Figure 5.13, what is the sequence of the polypeptide transcribed from the following mRNA strand: 5′ AUGGUGCAUAUUAGAUAC 3′?

- The polypeptide sequence would be: Met Val His Ile Arg Tyr.


**Figure 5.13** The genetic code. There are 64 possible mRNA codons (each consisting of three bases). To identify the amino acid coded by a particular codon, select the first base from the rows on the left, the second base from the columns along the top and the third base from the codons shown in the appropriate square. Note that AUG, the codon for Met, is also the 'start' codon. There are also three 'stop' codons. The abbreviated names and single letter code for the 20 amino acids found in proteins are listed (you do not need to remember these abbreviations, nor which codons correspond to which amino acids).

You should by now be beginning to see how certain changes (mutations) in the base sequence of a gene might affect the sequence of the encoded protein.

- How might such mutations arise?

  They may arise as a result of uncorrected errors during DNA replication, or as a result of unrepaired damage, such as deamination of bases, formation of thymine dimers or double-strand breaks.

A section of sequence that encodes a polypeptide sequence, or part of a polypeptide sequence is known as an **open reading frame (ORF)**. An ORF is therefore a series of triplet codons, uninterrupted by a stop codon. A mutation that disrupts an ORF may change the protein product. In the next section you will study how different types of gene mutation in coding, and also non-coding regions may lead to an effect on the function of a gene or its protein product.

## 5.6 The consequences of mutations

The rate at which mutations – permanent changes in DNA sequence – accumulate varies between different species, and even between different regions of DNA in the same species. It can be as low as one in every $10^{10}$ bp per genome per cell division, to as high as one in every 100 bp per genome per cell division (which is common in unicellular eukaryotes and bacteria).

The mutation rate can be estimated by observing the rate at which spontaneous changes arise in specific genes, either in a population of organisms, or in cells from the organism growing in culture in the laboratory. Application of both of these methods to an analysis of mutation rate in the mouse genome has suggested an uncorrected error frequency of only one base pair change in every $10^9$ bp for each cell generation. The mouse genome is about the same size as the human genome (about $3 \times 10^9$ bp), so there are only a couple of new errors in each newly divided cell.

Consequently, a typical mouse gene containing about $10^3$ protein coding base pairs would suffer a mutation once in about $10^6$ cell generations (i.e. $10^9$ bp / $10^3$ bp). A total of about $10^{12}$ cell divisions take place during the lifetime of a mouse. Thus, on average, every single gene would have acquired a mutation, somewhere in a mouse's body, on about $10^6$ (i.e. $10^{12} / 10^6$) occasions. The older the mouse is, the greater the total number of mutations that will have accumulated in its cells. This explains why the incidence of cancer in mammals increases with age: cancer results from the accumulation in a cell of several mutations that disrupt the function of genes that control cell division (Book 3, Chapter 1).

In eukaryotes, only those mutations that occur in the germ-line, the cells that ultimately form the gametes and therefore contribute to the next generation, can give rise to heritable gene variants. Chapter 4 looked at the patterns of inheritance of genes in a variety of biological systems. In all these cases, the phenotype of a particular individual is determined by combinations of gene variants known as alleles. But why are some alleles of a gene dominant to others? Why are some mutant alleles recessive to their wild-type counterpart, while others are dominant to the wild type? To answer these questions, you will now move on to look at the nature of mutations, and at the molecular consequences of different types of mutation.

## 5.6.1 Changes in single nucleotides

Gene mutations that result in a simple exchange of one type of base for another may have a very subtle effect on a gene product. In Section 5.4.2 you learnt about changes in individual bases in the DNA that are caused by spontaneous chemical modification of a base, for example, the deamination of cytosine to form uracil (Figure 5.7a), and depurination (loss of a guanine or adenine base).

- From your study of this chapter, how can abasic sites such as those derived from depurination be repaired?

  Abasic sites are repaired by enzymes that cleave the sugar–phosphate backbone either side of the abasic site. The resulting gap is filled in by the action of DNA polymerase and ligase. This is base excision repair.

If a uracil residue resulting from deamination is not recognised and removed before DNA replication occurs, at the second round of replication A would pair with T, and thus the original C:G pair would be replaced by a T:A pair (Figure 5.7b).

The second type of event to effectively bring about a base change would be the misincorporation of an incorrect nucleotide during replication. *Transitions* involve the substitution of one pyrimidine (C or T) by the other, or of one purine (A or G) by the other, while *transversions* involve the replacement of a pyrimidine by a purine or vice versa.

- Is the mutation C changed to U a transition or a transversion?

- It is a transition, since both cytosine and uracil are pyrimidines.

Base changes within an ORF can have a number of consequences for the protein product of the gene. They may cause substitution of one amino acid for another, known as a ***missense mutation*** (Figure 5.14a), or they may introduce a premature stop codon, which is known as a *nonsense mutation* (Figure 5.14b).

- What would be the consequence of a nonsense mutation for a protein encoded by the mutated gene?

  The stop codon would prevent translation of the full length of the ORF and the result would be a truncated protein product.

The more subtle change caused by a missense mutation might have a drastic effect on protein function if it changed an essential amino acid, for example in the active site of a protein (Book 2, Chapter 1). On the other hand, many base changes cause absolutely no change to the protein sequence. These are known as *silent mutations*.

- Looking at the genetic code in Figure 5.13, can you suggest why some single base changes within an ORF might have no effect on the encoded protein?

  Some amino acids are encoded by several codon sequences, which often differ only in the third base, so changes in the third base of a codon often don't alter the amino acid incorporated into the protein (Figure 5.14c).

Errors in DNA replication can also lead to the insertion or deletion of one or more nucleotides, as shown in Figure 5.14d and e. Such mutations can have a significant effect on the gene. Recall that nucleotides are decoded as triplet codons (groups of three nucleotides) during translation. Imagine that an extra nucleotide is inserted somewhere in an ORF encoding a section of protein. The consequence is that the register, or *reading frame*, after the insertion will be shifted along one nucleotide, so that the codons, and hence the amino acid sequence encoded, from that point onwards would be completely changed, and the structure of the polypeptide encoded by the gene would be significantly altered, with probable disruption of its function. Such mutations are known as ***frameshift mutations***.

- Would the insertion of three nucleotides into an ORF cause a shift in reading frame?

  No; because the genetic code is translated in triplets, insertions or deletions of multiples of three nucleotides do not cause a frameshift.

The mutation rate can be estimated by observing the rate at which spontaneous changes arise in specific genes, either in a population of organisms, or in cells from the organism growing in culture in the laboratory. Application of both of these methods to an analysis of mutation rate in the mouse genome has suggested an uncorrected error frequency of only one base pair change in every $10^9$ bp for each cell generation. The mouse genome is about the same size as the human genome (about $3 \times 10^9$ bp), so there are only a couple of new errors in each newly divided cell.

Consequently, a typical mouse gene containing about $10^3$ protein coding base pairs would suffer a mutation once in about $10^6$ cell generations (i.e. $10^9$ bp $10^3$ bp). A total of about $10^{12}$ cell divisions take place during the lifetime of a mouse. Thus, on average, every single gene would have acquired a mutation, somewhere in a mouse's body, on about $10^6$ (i.e. $10^{12} \div 10^6$) occasions. The older the mouse is, the greater the total number of mutations that will have accumulated in its cells. This explains why the incidence of cancer in mammals increases with age: cancer results from the accumulation in a cell of several mutations that disrupt the function of genes that control cell division (Book 3, Chapter 1).

In eukaryotes, only those mutations that occur in the germ-line, the cells that ultimately form the gametes and therefore contribute to the next generation, can give rise to heritable gene variants. Chapter 4 looked at the patterns of inheritance of genes in a variety of biological systems. In all these cases, the phenotype of a particular individual is determined by combinations of gene variants known as alleles. But why are some alleles of a gene dominant to others? Why are some mutant alleles recessive to their wild-type counterpart, while others are dominant to the wild type? To answer these questions, you will now move on to look at the nature of mutations, and at the molecular consequences of different types of mutation.

## 5.6.1 Changes in single nucleotides

Gene mutations that result in a simple exchange of one type of base for another may have a very subtle effect on a gene product. In Section 5.4.2 you learnt about changes in individual bases in the DNA that are caused by spontaneous chemical modification of a base, for example, the deamination of cytosine to form uracil (Figure 5.7a), and depurination (loss of a guanine or adenine base).

- From your study of this chapter, how can abasic sites such as those derived from depurination be repaired?

- Abasic sites are repaired by enzymes that cleave the sugar–phosphate backbone either side of the abasic site. The resulting gap is filled in by the action of DNA polymerase and ligase. This is base excision repair.

If a uracil residue resulting from deamination is not recognised and removed before DNA replication occurs, at the second round of replication A would pair with T, and thus the original C:G pair would be replaced by a T:A pair (Figure 5.7b).

The second type of event to effectively bring about a base change would be the misincorporation of an incorrect nucleotide during replication. *Transitions* involve the substitution of one pyrimidine (C or T) by the other, or of one purine (A or G) by the other, while *transversions* involve the replacement of a pyrimidine by a purine or vice versa.

- Is the mutation C changed to U a transition or a transversion?

- It is a transition, since both cytosine and uracil are pyrimidines.

Base changes within an ORF can have a number of consequences for the protein product of the gene. They may cause substitution of one amino acid for another, known as a *missense mutation* (Figure 5.14a), or they may introduce a premature stop codon, which is known as a *nonsense mutation* (Figure 5.14b).

- What would be the consequence of a nonsense mutation for a protein encoded by the mutated gene?

  The stop codon would prevent translation of the full length of the ORF and the result would be a truncated protein product.

The more subtle change caused by a missense mutation might have a drastic effect on protein function if it changed an essential amino acid, for example in the active site of a protein (Book 2, Chapter 1). On the other hand, many base changes cause absolutely no change to the protein sequence. These are known as *silent mutations*.

- Looking at the genetic code in Figure 5.13, can you suggest why some single base changes within an ORF might have no effect on the encoded protein?

  Some amino acids are encoded by several codon sequences, which often differ only in the third base, so changes in the third base of a codon often don't alter the amino acid incorporated into the protein (Figure 5.14c).

Errors in DNA replication can also lead to the insertion or deletion of one or more nucleotides, as shown in Figure 5.14d and e. Such mutations can have a significant effect on the gene. Recall that nucleotides are decoded as triplet codons (groups of three nucleotides) during translation. Imagine that an extra nucleotide is inserted somewhere in an ORF encoding a section of protein. The consequence is that the register, or *reading frame*, after the insertion will be shifted along one nucleotide, so that the codons, and hence the amino acid sequence encoded, from that point onwards would be completely changed, and the structure of the polypeptide encoded by the gene would be significantly altered, with probable disruption of its function. Such mutations are known as *frameshift mutations*.

- Would the insertion of three nucleotides into an ORF cause a shift in reading frame?

- No; because the genetic code is translated in triplets, insertions or deletions of multiples of three nucleotides do not cause a frameshift.

RNA polymerase or a regulatory protein) may reduce the amount of functional protein in the cell.

In addition, large chromosome rearrangements (Section 5.11) occasionally relocate the gene to so-called 'silent' chromosome regions that are unfavourable for transcription, such as heterochromatin (Section 3.4.3). Insertion of a large DNA fragment can also disrupt transcription of a gene and often leads to a null mutation; this is typical of mutations caused by the mobilisation of transposable elements (Section 5.8.2).

In contrast, mutant alleles that are dominant over wild type generally represent a change in function, or an increase in activity of the gene product. **Gain of function mutations** (sometimes referred to as *hypermorphs*) may be due to:

- increased activity of the gene product
- increased levels of transcription
- inappropriate patterns of gene expression, which may be temporal (i.e. when a gene is transcribed) or spatial (where a gene is transcribed); inappropriate gene expression can have drastic effects on the development of an organism
- large chromosomal rearrangements that move the gene's coding region into proximity with DNA transcriptional control sequences that normally activate another gene.

Very rarely, a mutation can lead to a gain of gene function that is completely different from the original function (referred to as a ***neomorph***). A new function may result from fusion of the coding regions of two different genes, or a sequence change that leads to the production of an aberrant protein with new activity.

It would be a mistake, however, to think that genetic mutation is an 'all or nothing' phenomenon – different mutations may lead to a subtly graded series of phenotypes. To return to the example of the *Drosophila* gene *white*, a great number of different mutant alleles have been characterised in the 100 years or so since the first mutant allele was identified. The phenotypes caused by these alleles range from the pure white eyes of *white* alleles that cannot express any functional protein at all to those with only a very slight effect on the normal brick-red eye colour.

## Summary of Sections 5.4 to 5.6

- Uncorrected changes in DNA lead to mutations, permanent changes in DNA sequence that will be copied and passed to daughter cells when the cell divides. Errors resulting from replication and DNA damage are repaired by a variety of DNA repair pathways, many of which rely on the undamaged DNA strand to provide a template for an accurate repair.
- Groups of three nucleotides (codons) in the DNA or mRNA sequence (the genetic code) specify the order in which amino acids will be added to the polypeptide chain when an mRNA sequence is translated.
- Mutations that change, insert or delete bases may affect the amino acid sequence of an encoded protein, or may alter non-coding sequences that

affect gene expression. In eukaryotes, most mutations have no effect because they occur in regions of the genome that do not encode (or regulate) proteins or other functional gene products.

- Mutations may create recessive (usually complete or partial loss of gene function) or dominant (usually gain of function or novel function) alleles of a gene.

## Classical genetics in prokaryotes

The early research in classical genetics was driven largely by work on multicellular eukaryotes like *Drosophila*. By contrast, the early advances in the understanding of the molecular nature of genes and their regulation were made through the study of prokaryotic organisms. Principal among these were *E. coli* and the bacteriophage (viruses of bacteria) that infect it. The discussion of gene and genome structure will therefore begin by considering prokaryotic genomes.

### 5.7.1 Prokaryotic gene structure

The genes of prokaryotes are simpler in structure than those of eukaryotes. In general, the nucleotide sequence that comprises the protein-coding region of a prokaryotic gene is an uninterrupted series of codons, beginning with a codon for methionine (ATG) and terminating with one of the stop or termination codons (TAA, TAG or TGA). Translation of the mRNA begins at the methionine codon, and continues until the stop codon is reached. there are no other 'punctuation' signals within the prokaryotic ORF.

The ORF is typically preceded by a regulatory region containing DNA sequences that regulate the level of gene expression, and which you will study in more detail in Chapter 6. The regulatory region of a prokaryotic gene includes DNA sequences which provide the binding site for the enzyme RNA polymerase, which synthesises mRNA, and also sequences to which regulatory proteins bind. These regulatory proteins interact with the RNA polymerase to increase or decrease the efficiency of gene transcription, and are often exquisitely sensitive to the cell's physiology. The ability of regulatory proteins to bind DNA often depends on the availability of nutrients in the environment, or molecules required in metabolic pathways within the cell. This allows the cell to rapidly alter gene expression in response to its requirements.

Genes in prokaryotes are frequently arranged in units called **operons**. An operon is a group of genes with related function that are under the control of a single regulatory DNA region (Figure 5.15), the genes are transcribed together as a single mRNA called a **polycistronic transcript** (from cistron, an alternative word for gene), from which each of the individual proteins are then translated. This contrasts with eukaryotes, in which each individual gene usually has its own regulatory region.

One classic example of a prokaryotic operon, and one that is particularly well understood is the *lac* operon in *E. coli*. The *lac* operon includes three ORFs that are transcribed as a single polycistronic mRNA. The three ORFs of the *lac* operon encode enzymes that are related to the use of the sugar lactose as a

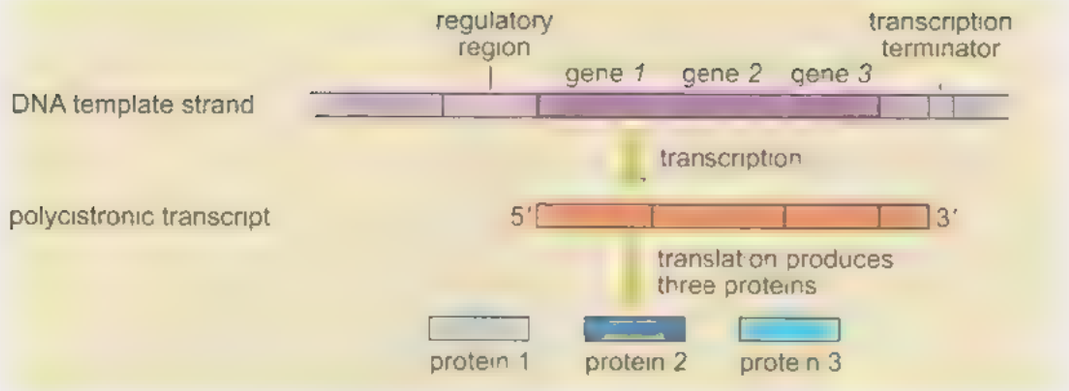



**Figure 5.15** A bacterial operon, a group of genes controlled by one regulatory region. The resulting polycistronic transcript shown here is translated to produce three different gene products (proteins).

carbon source. *lacZ* encodes a β-galactosidase (an enzyme that converts the disaccharide lactose into the monosaccharides glucose and galactose), *lacY* encodes a protein required for the import of lactose into the cell, while *lacA* encodes an enzyme that is not strictly essential for lactose metabolism. The *lac* regulatory region ensures that all three genes are expressed only when lactose is available. You will look at the control of the *lac* operon in more detail in Chapter 6.

## 5.7.2 Prokaryotic genomes and plasmids

Most prokaryotes have a single circular chromosome, that of an *E. coli* bacterium is about $4.6 \times 10^6$ bp in total. While prokaryotic DNA is not packaged with protein to the same extent as eukaryotic DNA (see below), it is not simply a loose circle of DNA. In fact, the prokaryotic chromosome is compacted by twisting of the DNA duplex, rather like coiling up a cable. Enzymes called topoisomerases (which you have already encountered in the discussion of DNA replication in Section 5.3 above) act to add or remove these twists. The torsional stress of coiling the circular DNA molecule causes it to wind up on itself (Figure 5.16) which is referred to as **supercoiling**.

The genome of a typical prokaryote has very little 'non-gene' DNA; it consists of genes and operons separated by very short stretches of DNA. This is in sharp contrast with the much larger genomes of eukaryotes, where in part the large size is due to the much wider spacing between genes.

In addition to the genomic chromosomal DNA described above, most prokaryotes have separate *extrachromosomal* elements, known as **plasmids**. Plasmids are small circular genetic elements (also known as **episomes**) that can replicate independently from the genomic chromosome. In many ways, plasmids function as small independent genomes: they generally have a single origin of replication, and their replication may be linked to the cell's division cycle. Plasmids often carry a small number of genes associated with a specific set of functions. For example, plasmids often encode proteins that facilitate their own movement from cell to cell. F plasmids (named F for fertility factor) encode genes required for the bacterium to form sex pili – cell surface

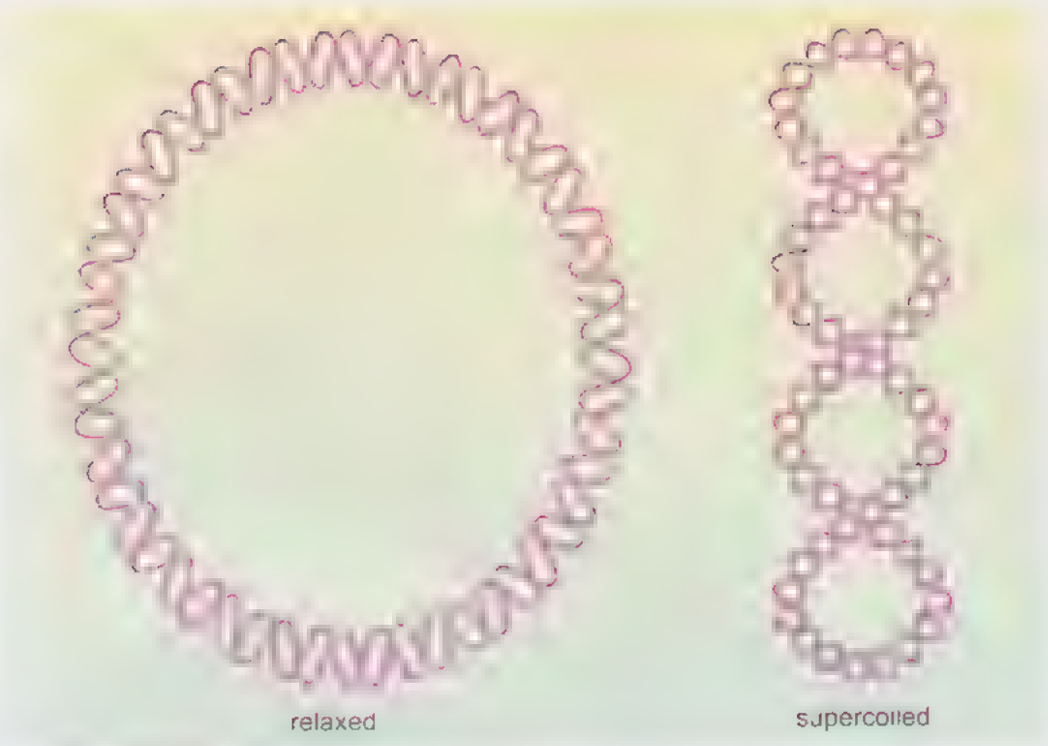



**Figure 5.16** Bacterial chromosomes are compacted by supercoiling. Typically, prokaryotic genomes are circular DNA molecules which are supercoiled by the introduction of twists into the double helix.

structures through which the plasmids, and in some cases chromosomal DNA, are transmitted from one cell to another during *conjugation*. This is one of a number of mechanisms of *horizontal gene transfer* (Section 5.7.5), which is a major source of microbial diversity in the wild.

The spread of plasmids through bacterial populations can also have an impact on human health. Some plasmids carry genes encoding toxins; the enterohaemorrhagic *E. coli* O157:H7 carries the pO157 virulence plasmid which encodes a potent toxin that causes diarrhoea. Plasmids may also encode virulence factors which increase the infectivity or harmful effects of pathogens, for example proteins that help bacteria to adhere to other cells, or to penetrate cell membranes. Plasmids have a role in the spread of antibiotic resistance genes in bacteria, perhaps most notoriously the genes for antibiotic-inactivating enzymes such as the β-lactamases which hydrolyse penicillin-like antibiotics, preventing them from binding to their target bacteria. Plasmids have proven invaluable in biological research as, among other things, they made possible the technique of gene cloning (Box 5.2).

## Box 5.2 Gene cloning

Gene cloning is a set of techniques by which individual genes can be isolated, manipulated and transferred between species. These techniques began to be developed during the 1970s as a way of studying gene sequence and function. Gene cloning depends on the universality of DNA. DNA molecules behave similarly irrespective of the source

organism, so a DNA molecule derived from one species has the capacity to be replicated, transcribed and translated in a second species. In order to clone genes, three technical advances were necessary. The first requirement was a reproducible means of cleaving DNA into defined sections. This was enabled by the discovery of **restriction endonucleases**, enzymes that catalyse the cleavage of isolated DNA molecules at specific sequences *in vitro*. The second advance was the development of techniques for joining together sections of DNA. Joining DNA fragments *in vitro* is accomplished by the action of DNA ligase.

- What processes in living cells require DNA ligase?

  Joining Okazaki fragments during lagging-strand DNA synthesis, and several of the mechanisms that repair DNA (Sections 5.3 and 5.4, respectively).

Finally, a means of propagating multiple copies of a specific DNA fragment was required. This can be accomplished by inserting the DNA fragment into a DNA molecule that has the capability to replicate independently within an appropriate host cell, and can be engineered *in vitro* to possess features that facilitate this function. Such molecule are generally referred to as **cloning vectors** and a large range of vectors suitable for propagation in different types of cells have been developed from naturally occurring plasmids, and also bacteriophage and virus genomes. The process of gene cloning using a bacterial plasmid vector is summarised here.

Restriction endonucleases (often referred to as *restriction enzymes*) cleave DNA strands at specific sequences. The majority of restriction enzymes are isolated from prokaryotes, where their natural function is to cleave any foreign DNA, such as bacteriophage DNA, that enters the cell, thereby evading infection. There are several different types of restriction enzyme, but the most useful in gene cloning are those that recognise a specific short DNA target sequence, and cut both strands of the DNA duplex. The target DNA sequence is usually palindromic (that is, it has the same sequence read in both directions), and in the case of the most widely used restriction enzymes, the two DNA strands are cleaved in a staggered fashion (Figure 5.17).

The important thing to note from Figure 5.17 is that each of the enzymes shown cleaves DNA to yield DNA fragments that have ends, or termini, with overhanging single strands of DNA. Furthermore, each of these enzymes always yields DNA with single-stranded ends of the same sequence, which, because of the palindromic nature of the target sequence, are complementary. Such complementary termini are called *cohesive* (or 'sticky') *ends* because they will base-pair with each other, and can be permanently rejoined *in vitro* using the enzyme DNA ligase to link the sugar–phosphate backbone of the fragments. It is the specificity afforded by base pairing between complementary termini, and the ease of joining then back together, that is the key to gene cloning techniques. Essentially, plasmid vector DNA is cleaved with a restriction enzyme to open up the circle, and mixed with the DNA fragment of

interest – derived by cleavage with the same restriction enzyme. By incubating this mixture in the presence of DNA ligase, **recombinant** DNA molecules in which the DNA fragment of interest has been inserted into the circular plasmid can be obtained (Figure 5.18).

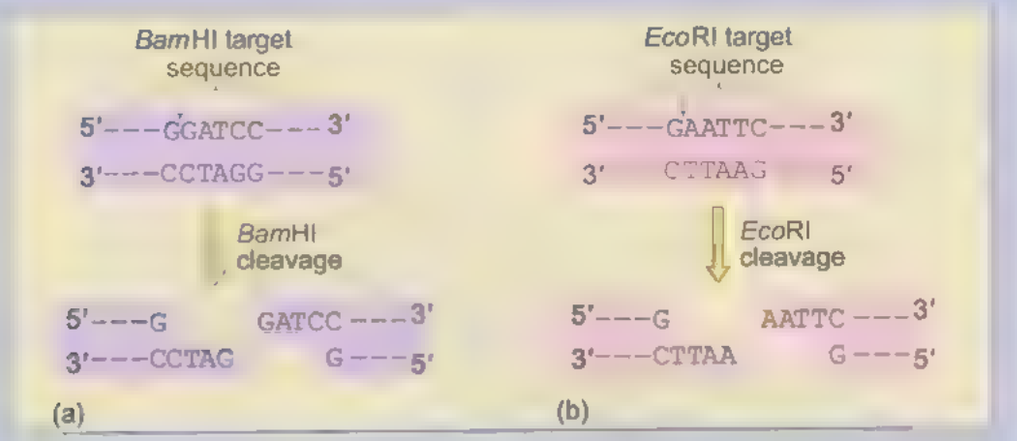
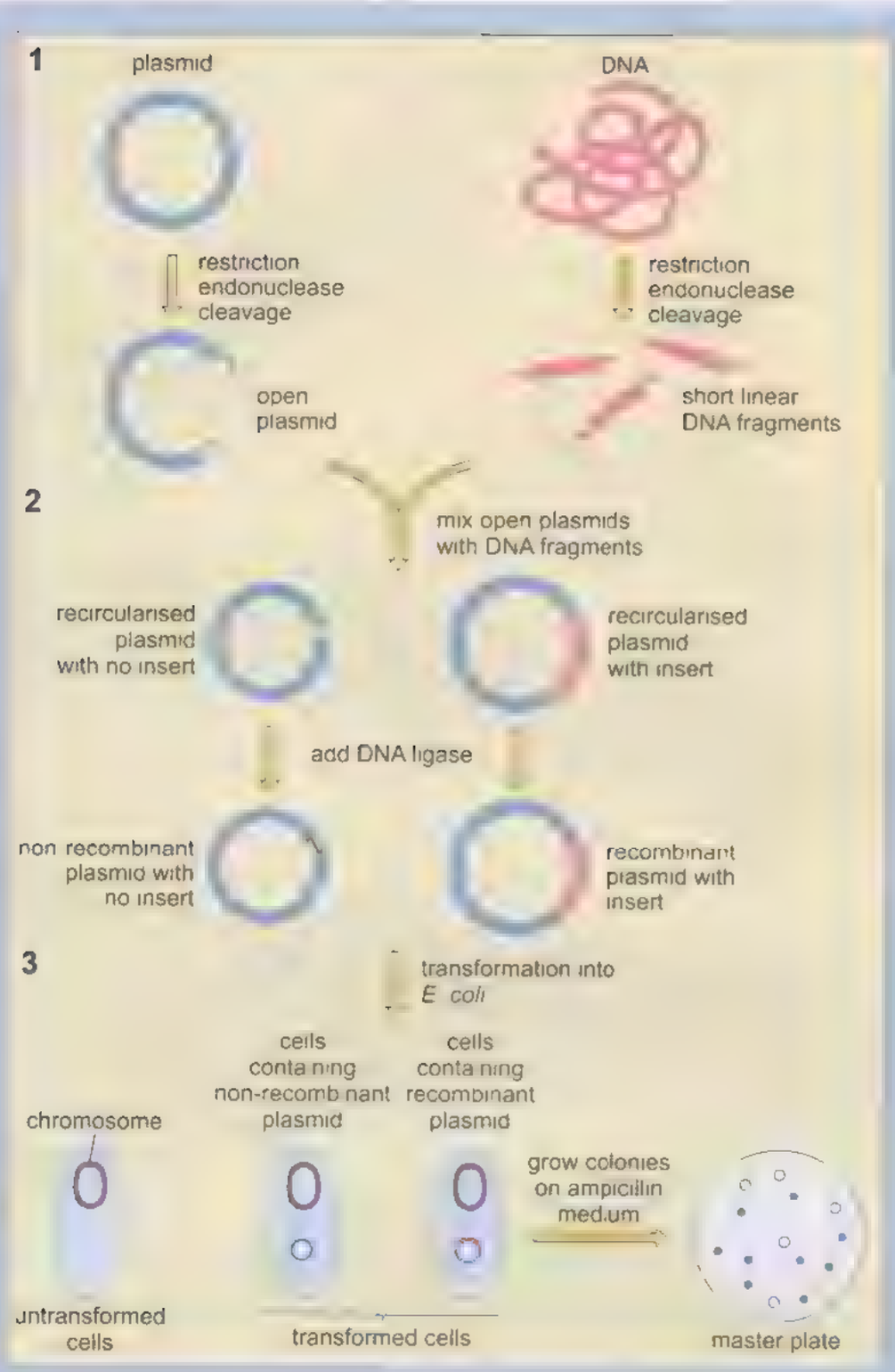



**Figure 5.17** Restriction endonucleases used for gene cloning can produce protruding single-stranded 'sticky' termini for example, (a) *Bam*HI and (b) *Eco*RI. The blue arrows indicate where the DNA strands are cleaved in the specific target sequences.

The final stage in DNA cloning is the propagation of the DNA fragment of interest. Typically, this is done within cells of the bacterium *E. coli*, and takes advantage of the plasmid vector's ability to replicate within a bacterial cell. In the laboratory, living *E. coli* cells can be induced by certain treatments to take up the recombinant plasmid DNA in a process known as *transformation*; such cells are referred to as *competent* cells (Section 5.7.5). Once inside the cell, the recombinant plasmid DNA replicates and provides a source of large quantities of the DNA fragment of interest.

**Figure 5.18** Flowchart of DNA cloning. (1) A cloning vector, in this case a plasmid (such as that shown in Figure 5.19), is cleaved with a restriction endonuclease, as is a preparation of DNA to be cloned. (2) The two DNA preparations are mixed to allow the cohesive termini of the DNA fragments to anneal. DNA ligase rejoins the sugar–phosphate backbone of the DNA molecules. Note that some re-ligated plasmid molecules will not contain the DNA of interest. (3) *E. coli* cells that have been pretreated to enable them to take up DNA are transformed with the ligated DNA, and propagated on agar plates containing the antibiotic ampicillin. Only the cells that contain plasmid can grow. If the agar plates are also supplemented with an appropriate chromogenic substrate for β-galactosidase (see text below), cells containing plasmids in which *lacZ* is interrupted by the insertion of a DNA segment will form white colonies, while cells containing plasmids with no insert (and therefore a functional *lacZ*) will form blue colonies.




Commonly used plasmid cloning vectors, like the one shown in Figure 5.19, are engineered to introduce a number of useful features.

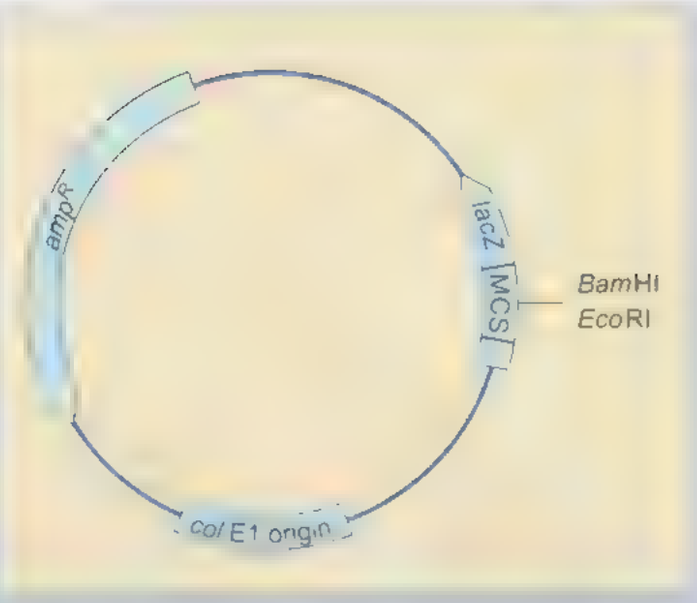



**Figure 5.19** A typical plasmid cloning vector. This circular molecule contains a replication origin (*col*E1) and a gene for ampicillin resistance ($amp^R$). The block marked MCS (multiple cloning site) contains the recognition sequence for several restriction enzymes. The MCS is located within the *E. coli lacZ* gene.

These include:

- *An origin of replication.* The example in Figure 5.19 has a *col*E1 origin of replication (derived from a naturally occurring *E. coli* plasmid).
- *An antibiotic resistance gene.* Because transformation is inefficient, only a small proportion of cells will take up the plasmid, so a means of *selecting* those cells that have successfully taken up the recombinant plasmid is needed. For example, by including a gene that confers resistance to the antibiotic ampicillin in the plasmid vector ($amp^R$ in Figure 5.19), only those cells that have successfully taken up the plasmid will grow on selective medium containing ampicillin, while cells that do not harbour the plasmid vector will fail to grow (Figure 5.18).
- *A choice of several different restriction enzyme sites at which DNA may be inserted* (as shown in Figure 5.19). To enhance versatility, most modern vectors can accept DNA obtained by cleavage with any of a number of different restriction enzymes. The advent of the polymerase chain reaction (Box 5.1) has in many cases simplified the isolation of specific DNA sequences for gene cloning.
- *A means of distinguishing recombinant plasmids, which contain an inserted DNA fragment, from recircularised 'empty' vector plasmid.* For example, the plasmid shown in Figure 5.19 bears the restriction enzyme cloning sites within a copy of the *E. coli lacZ* gene that encodes β-galactosidase (Section 5.7.1 above). When no DNA is inserted in the vector plasmid, the *lacZ* gene is expressed, and the cell carrying the plasmid will synthesise β-galactosidase enzyme,

which can be visualised by spreading the cells on agar growth medium including a substrate that the enzyme converts from colourless to blue (such a substrate is known as a *chromogenic substrate*). The bacterial colony that grows from the cell will therefore appear blue (Figure 5.18). When a DNA fragment is inserted into the plasmid, however, it disrupts the *lacZ* gene sequence, such that the gene is no longer functional and cannot catalyse the change in the chromogenic substrate, so the bacterial colonies carrying recombinant plasmids are white in colour.

This overview of DNA cloning is necessarily brief, and does not include the many techniques and vector systems that have been developed to identify specific DNA sequences within a large collection of DNA fragments, or to express recombinant proteins in different organisms.

## 5.7.3 Prokaryotic DNA replication

Earlier in this chapter, you learnt how DNA molecules were replicated. As you might imagine for a process that is so vitally important, the enzymes and other proteins involved in DNA replication are well *conserved*, that is, they have a very similar amino acid sequence in most species, from prokaryotes to eukaryotes. In general, prokaryotic chromosomes have a single point at which DNA replication is initiated, known as the origin of replication (Section 5.3). Replication proceeds bidirectionally from the origin of replication, as shown in Figure 5.20a. The initiation of replication is tightly controlled: it is linked to cell size and occurs once per cell division, although in rapidly dividing bacterial cells, a second round of replication may be initiated before the first is complete.

Plasmids in Gram-negative bacteria usually replicate bidirectionally in much the same way as a bacterial chromosome (Figure 5.20a), while those in Gram-positive bacteria usually replicate by another mechanism called rolling circle replication (Figure 5.20b). Linkage of plasmid replication to the host cell's cell division cycle is not always very strict, so there may be quite wide variation in copy number present in cells. Indeed, many plasmids used for gene cloning in the laboratory have been deliberately selected to have a very high copy number, in order to maximise the yield of recombinant DNA or of the recombinant proteins encoded by the plasmid.

## 5.7.4 Prokaryotic genomes are similar to eukaryotic organelle genomes

In Section 1.2.4, you encountered the endosymbiotic theory of the origin of mitochondria and chloroplasts, the organelles responsible for ATP synthesis and photosynthesis, respectively, in eukaryotes. Alone among the extranuclear organelles, mitochondria and chloroplasts possess their own genomes, which resemble prokaryotic genomes in that they are small and circular (Figure 5.21) and the DNA is not packaged with proteins in the same way as a typical eukaryotic chromosome.

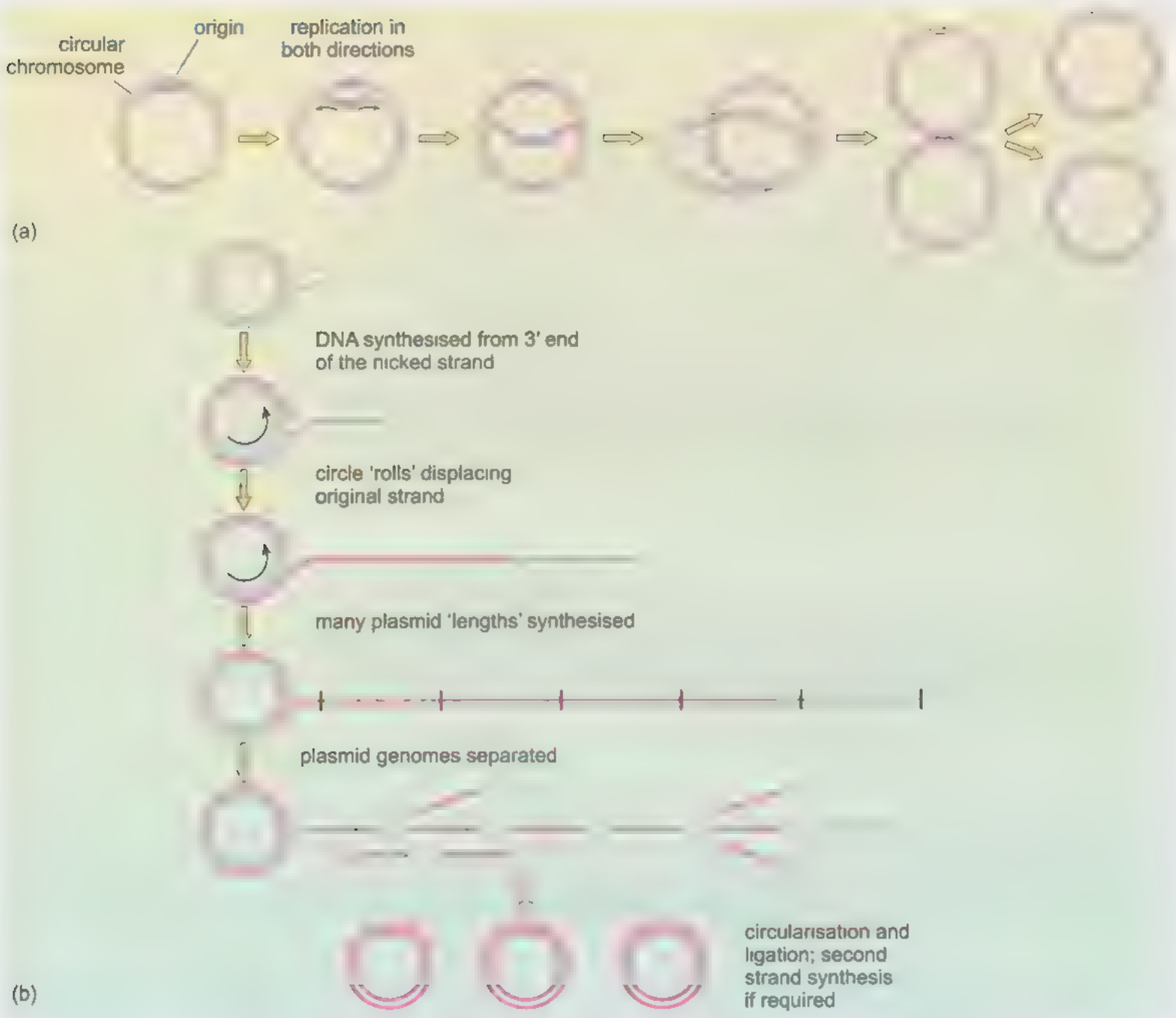



**Figure 5.20** Replication in prokaryotes (a) Bidirectional replication of a circular prokaryotic genome. Two replication forks proceed in opposite directions from a single origin of replication, and converge at a point opposite the origin. Many plasmids also replicate by this mechanism (b) The rolling circle mechanism of plasmid replication, typical of Gram-positive bacteria. Replication begins at the origin where one strand is nicked, providing a starting point for replication of the new DNA strand around the circle. Many copies of the template are made as one long strand and cut into individual genomes which then recircularise and provide a template for synthesis of a second DNA strand to give new double-stranded plasmids.

Mitochondrial and chloroplast genomes are substantially smaller than that of a 'typical' prokaryote such as *E. coli*. In fact, the great majority of the proteins in mitochondria and chloroplasts are actually encoded by the nuclear genome and imported into the organelle (Section 3.4); only a few are encoded in the organelle genome. The mitochondrial genome is particularly compact (Figure 5.21), with virtually no non-coding DNA (other than in the region

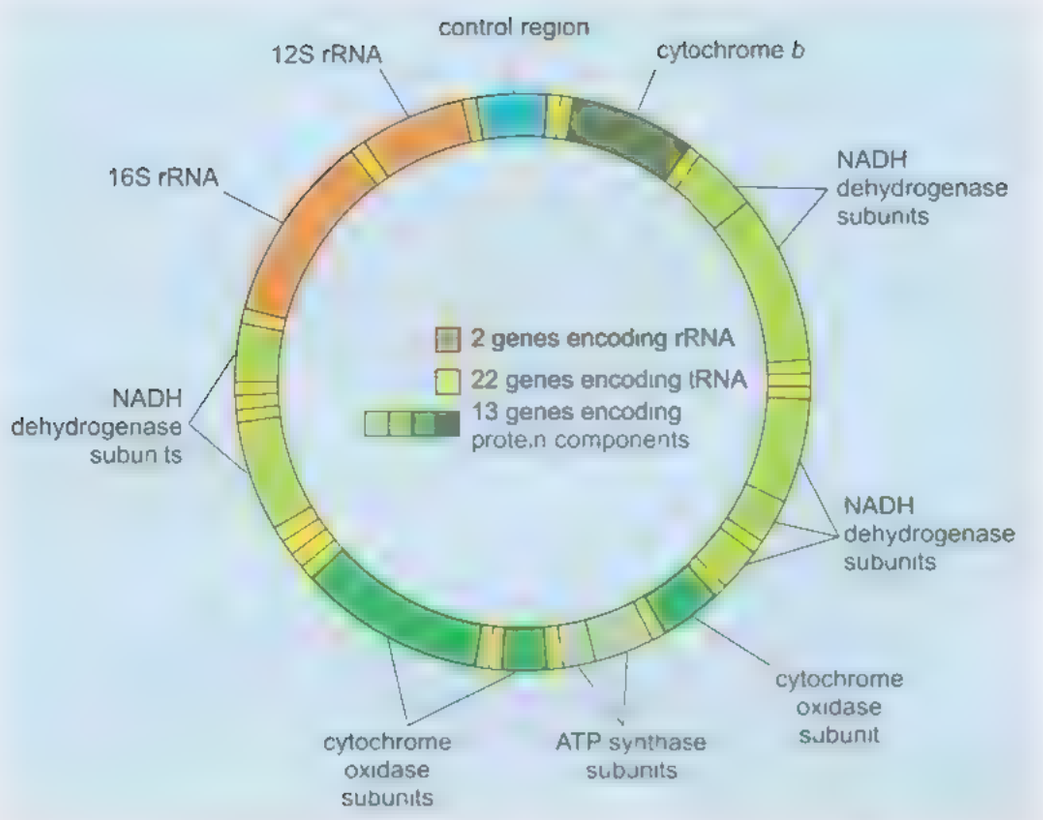



**Figure 5.21** The human mitochondrial genome is about $1.6 \times 10^4$ base pairs, and contains 37 genes encoding some proteins, transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs).

containing the replication origin) and in several cases, the open reading frame of one gene overlaps with that of the next. The mitochondrial genome encodes mitochondrial transfer RNAs, ribosomal RNAs, and a number of protein components of the electron transport chain required for ATP synthesis (Book 2, Chapter 3).

The genetic code shown in Figure 5.13 is universal for the nuclear genomes of all taxonomic groups. However, mitochondrial genomes deviate from this universal code. These differences mostly affect the start and stop codons (Table 5.1).

**Table 5.1** Differences between the 'universal' genetic code and the mammalian mitochondrial genetic code.

| | Encoded amino acid | |
|---|---|---|
| **Codon** | **Universal genetic code** | **Mitochondrial genetic code** |
| UGA | stop | tryptophan |
| AGA, AGG | arginine | stop |
| AUA | isoleucine | methionine |

## 5.7.5 Prokaryote genetics

Prokaryotes pass on their genes to their daughter cells through asexual binary fission, and the daughter cells are genetically identical *clones* of the parent cell (Activity 4.1). Bacteria and other prokaryotes do not have an equivalent to the sexual reproduction of eukaryotes (involving the production of gametes which fuse to produce offspring with new combinations of genetic material, Section 4.3.2). However, there are three possible mechanisms that allow **horizontal gene transfer** between two prokaryotic cells and which contribute to variation in prokaryotic populations.

### Conjugation

During conjugation, one bacterium extends a tubular structure known as the sex pilus towards another cell (Figure 5.22), and initiates the transfer of DNA through the sex pilus to the recipient cell. In *E. coli*, conjugation requires the cell to contain the F plasmid. F plasmids are quite variable in structure, but typically contain an origin of replication, an origin at which DNA transfer is initiated (known as *oriT*) and a set of *tra* genes that are required to form the sex pilus and initiate DNA transfer. ***Escherichia coli* strains containing F** plasmids are referred to as $F^+$ strains and those without as $F^-$ strains. During conjugation, a single strand of DNA is copied from the $F^+$ plasmid and passes through the sex pilus to the recipient $F^-$ cell, which therefore acquires a copy of the F plasmid (Figure 5.23a).

**Figure 5.22** This electron micrograph shows three bacterial cells during conjugation. Sex pili, through which DNA is transferred from one cell to another, are clearly visible.

In some bacterial strains, F plasmids can insert or *integrate* into the bacterial chromosome (Figure 5.23b). The integrated plasmid's *oriT* retains its activity, and upon conjugation it leads transfer of the bacterial chromosome through the sex pilus to the recipient cell where recombination can result in incorporation of donor strain genes into the chromosome of the recipient cell. Strains in which an F plasmid has become integrated in the chromosome are known as Hfr (high frequency of recombination) strains. The amount of chromosomal DNA that is transferred during conjugation depends on how long transfer can be maintained between the bacteria: sex pili are quite fragile and transfer is usually terminated by physical breakage of the pilus.

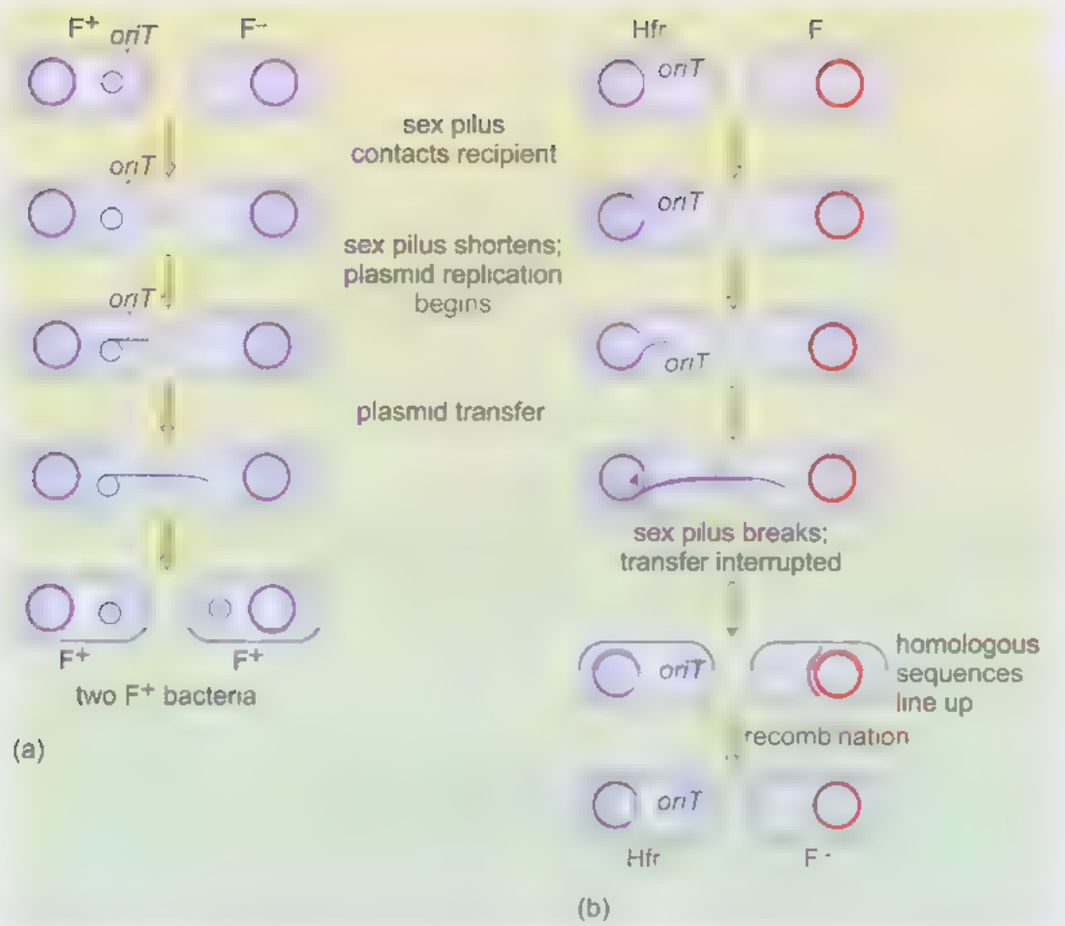



**Figure 5.23** Conjugation in *E. coli*. (a) Cells carrying F plasmids form sex pili through which a copy of the F plasmid may be transferred to cells lacking F plasmids (F⁻). (b) Hfr strains contain an F plasmid that has become integrated within the chromosome. During conjugation, DNA transfer initiates at the F plasmid's *oriT* and because the plasmid is integrated in the chromosome, the transfer includes all or part of the bacterial chromosome.

The transfer of chromosomal DNA by conjugation with an Hfr strain can be exploited in the laboratory to generate genetic recombination maps of the bacterial chromosome by 'interrupted mating' experiments (Figure 5.24). Cell suspensions from two cultures, one an Hfr strain and the other an F⁻ strain (with different genotypes) are mixed and left to conjugate. Samples are withdrawn at several time points, shaken vigorously to disrupt the sex pili and terminate conjugation, and plated on solid media. The recipient F⁻ cell may, for example, be mutant for several genes that encode enzymes required for the synthesis of essential substances (such as amino acids or nucleotides), so the cell is unable to grow on a 'minimal' medium lacking those substances (such mutants are known as *auxotrophs*). Conjugation with an Hfr strain that is wild type for these genes can transfer functional gene alleles to the F⁻ recipient cell by recombination into the recipient's genome, restoring these activities. The cells resulting from conjugation are then able to grow on minimal medium. The map is compiled by noting the time point at which transfer of wild type alleles from the Hfr strain restores the phenotype of the F⁻ recipient. It is for this reason that map positions in the *E. coli* genetic map are given in units of minutes. This type of recombination mapping in prokaryotes has, like the

recombination maps in eukaryotes (Section 4.5), now largely been superseded by DNA sequence analysis.

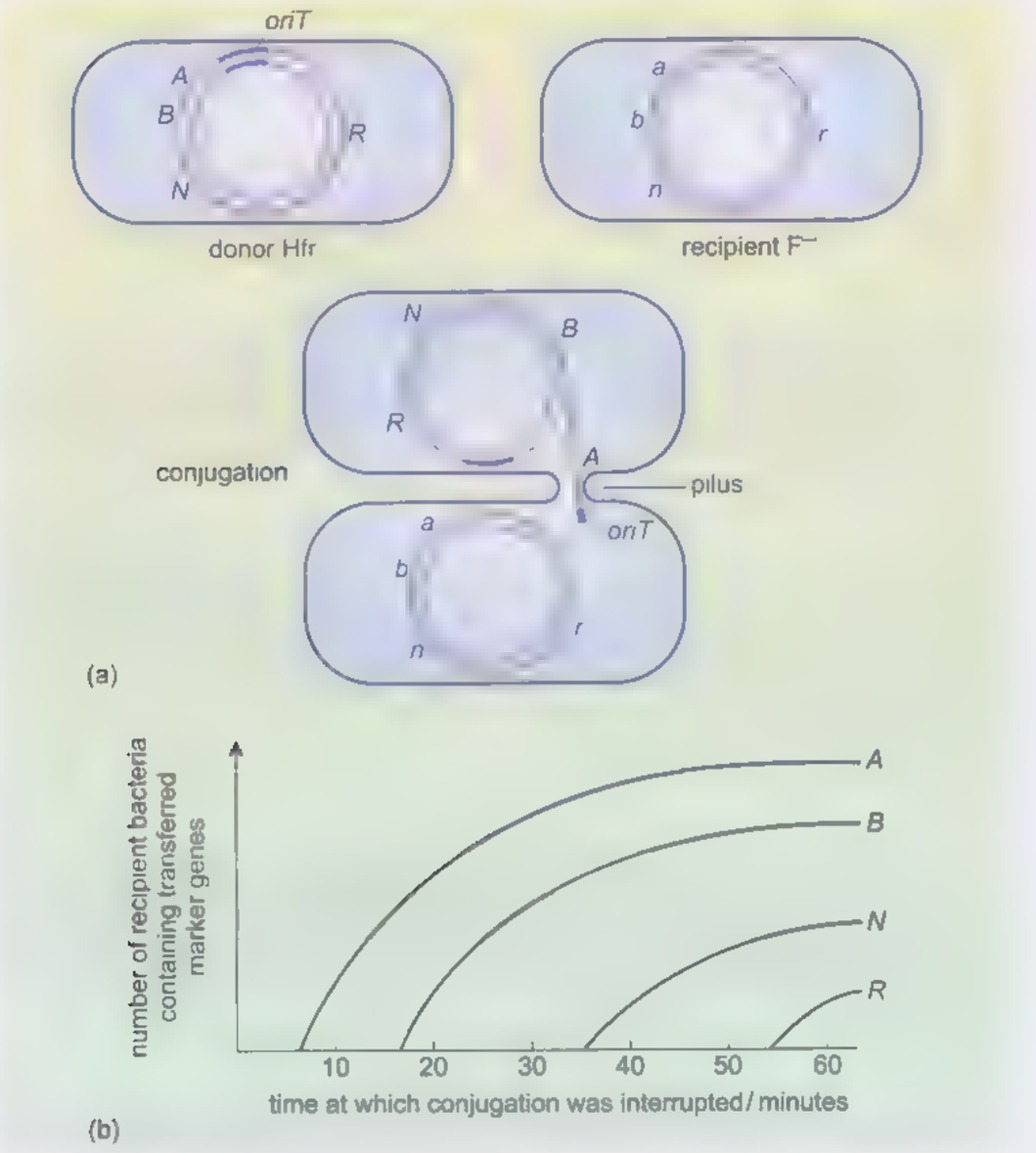



**Figure 5.24** Genetic mapping in *E. coli* using interrupted-mating conjugation. (a) The Hfr donor strain (genotype *A*, *B*, *N*, *R*) can conjugate with a recipient F cell (genotype *a*, *b*, *n*, *r*), during which the F plasmid integrated in the bacterial genome of the Hfr strain initiates the transfer of chromosomal DNA. Transfer continues until the sex pilus breaks (usually by physical shearing). (b) The appearance of new genotypes in the recipient depends on the gene distance from the F plasmid insertion site. In this example, gene *A* transfers before *B*, which in turn transfers before *N*, and *R* transfers last.

## Transformation

Bacteria are also able to take up 'naked' DNA from their surroundings and incorporate it into their genome by genetic recombination, a process known as **transformation**. This process occurs naturally in some bacteria, such as *Streptococcus pneumoniae* and *Haemophilus influenzae*, but most, including

*E. coli*, will only take up DNA if they are treated with metal ions such as calcium, which increase the permeability of their outer membrane; cells treated in this way are said to be *competent* to take up DNA. Other techniques that alter the permeability of cell membranes include electroporation, in which cells are exposed to pulsed electric fields. Artificially induced transformation is much less efficient than natural transformation, but if sufficiently high concentrations of DNA are used, it works adequately for experimental purposes.

Transformation can also be exploited for gene mapping. Competent cells will take up randomly sheared genomic DNA fragments, and if two genes are located near to each other on the bacterial chromosome, it is likely that some DNA fragments will contain both genes. The frequency of cotransformation for two genes that are close together would therefore be considerably higher than for two genes that are far apart. By conducting many such experiments, a genetic map based on cotransformation frequencies may be assembled. Transformation has, however, found its greatest use in gene cloning (Box 5.2), as a way of introducing recombinant plasmid DNA into host *E. coli* cells.

### Transduction

**Transduction** is the transfer of DNA from one cell to another mediated by a virus. When bacteriophage infect a bacterial cell, they use the host cell's replication, transcription, and translation machinery to make new viral genomes and viral coat proteins. The viral genomes are packaged into a *capsid* of viral proteins to form new infectious particles (Section 5.9). Some bacteriophage are able to integrate their genome into the host chromosome, and there remain dormant for some time (this is called *lysogeny*), but when the virus is activated to start proliferating, it excises from the bacterial genome, often accidentally carrying with it sections of the bacterial chromosome, which become copied and packaged into the new virus particles. This type of gene transfer has contributed to the transfer of antibiotic resistance and virulence factors between bacterial strains.

## Summary of Section 5.7

- The typical prokaryotic gene includes a continuous protein-coding open reading frame (ORF), with associated regulatory sequences to which RNA polymerase and various regulatory proteins bind.
- The genes in prokaryotes are often grouped in operons, which enable coordinately regulated expression of genes with related function.
- Typically, prokaryotic genomes are small and circular. Replication of the prokaryotic genome is initiated from a single origin of replication.
- Chloroplast and mitochondrial genomes share features with prokaryotic genomes (they are circular, and not compacted into chromatin), which is thought to reflect their origins in endosymbiosis.
- Plasmids are small circular DNA molecules that can replicate independently of the genomic DNA.

- Prokaryotic genomic DNA can undergo horizontal gene transfer between cells by the mechanisms of: conjugation (involving physical contact); transformation (involving transfer of naked DNA); or transduction (involving bacteriophage). These mechanisms contribute to variation in prokaryote populations and also the spread of antibiotic resistance and virulence factors.

## [illegible]

In contrast to prokaryotes, the genes and genomes of eukaryotes are generally far larger and more complex. In the early days of molecular biology, the prevailing expectation was that eukaryotic genome size would reflect the number of genes. However, it was quickly realised that genome size did not correlate with perceived organismal complexity; for example, many amphibians and flowering plants have genomes significantly larger than those of other multicellular organisms such as reptiles and mammals (Figure 5.25).

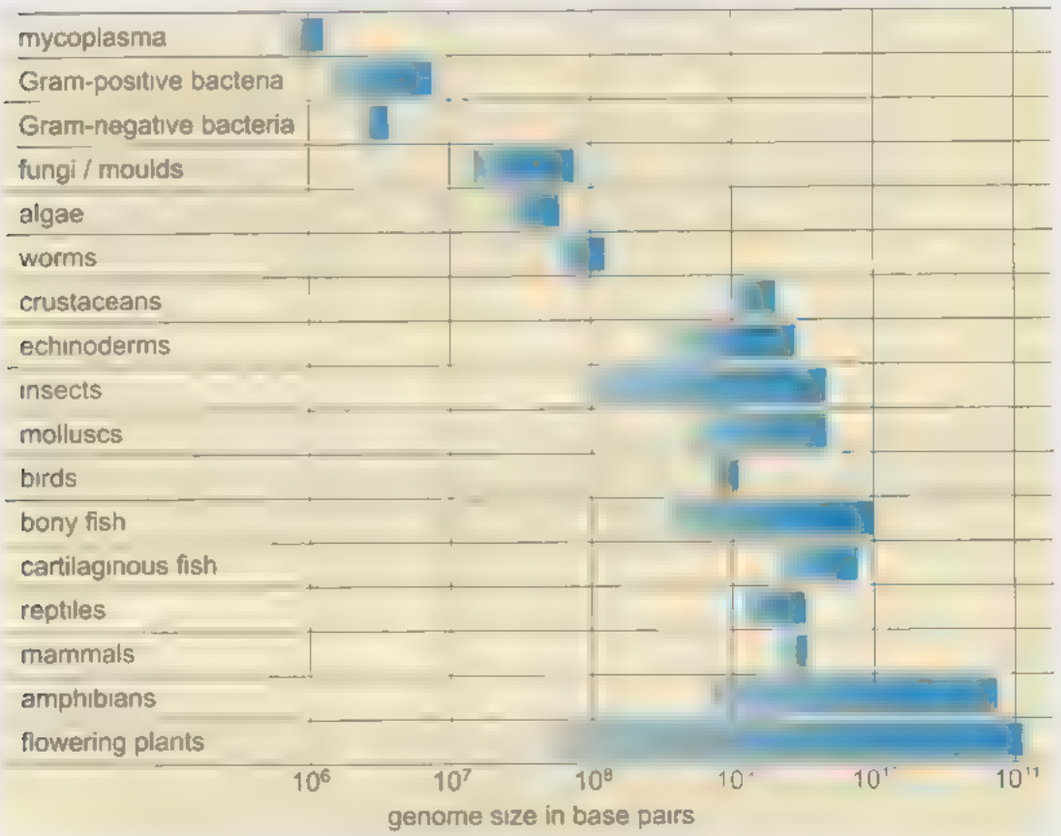



**Figure 5.25** A chart showing the variation in genome sizes across several taxonomic groups. Note the x-axis is a logarithmic scale (each interval is ten times greater than the previous one). Most mammalian genomes are around $3 \times 10^9$ bp.

The extensive variation in eukaryotic genome size was referred to as the C-value paradox (C being the amount of DNA in the genome). Subsequently the data from a number of eukaryotic genome sequencing projects revealed that gene numbers did not correlate with genome size, and were usually somewhat lower than expected. Early predictions of the number of human

genes were in the hundreds of thousands or even millions, but the completed genome sequence suggests that there are only around 20 000–25 000 genes, a rather surprising figure given the rather large $3 \times 10^9$ bp haploid genome size! In comparison, the fruit fly *Drosophila* has around 14 000 genes in its $1.7 \times 10^8$ bp haploid genome, while the worm *C. elegans* has around 19 000 genes in its $10^8$ bp genome. In fact, the human genome isn't unusually large. Figure 5.25 shows the range of genome sizes in some different taxonomic groups. What then is the explanation for the large size of eukaryotic genomes in comparison with the number of genes?

## 5.8.1 Eukaryotic genes are more complex than those of prokaryotes

The first point to consider is size of a typical eukaryotic gene. A schematic view of gene structure in eukaryotes is shown in Figure 5.26. In general, eukaryotic genes are not organised into operons; instead, each gene has its own regulatory DNA sequences. Another striking difference from prokaryotic genes is that the protein-coding region of a eukaryotic gene is usually interrupted by intervening non-coding sequences called **introns**, the coding sections are known as **exons**. Introns are removed by a post-transcriptional process known as **mRNA splicing**, to yield the mature messenger RNA (mRNA) in which the exons are contiguous and form the ORF required for protein translation (Chapter 6).

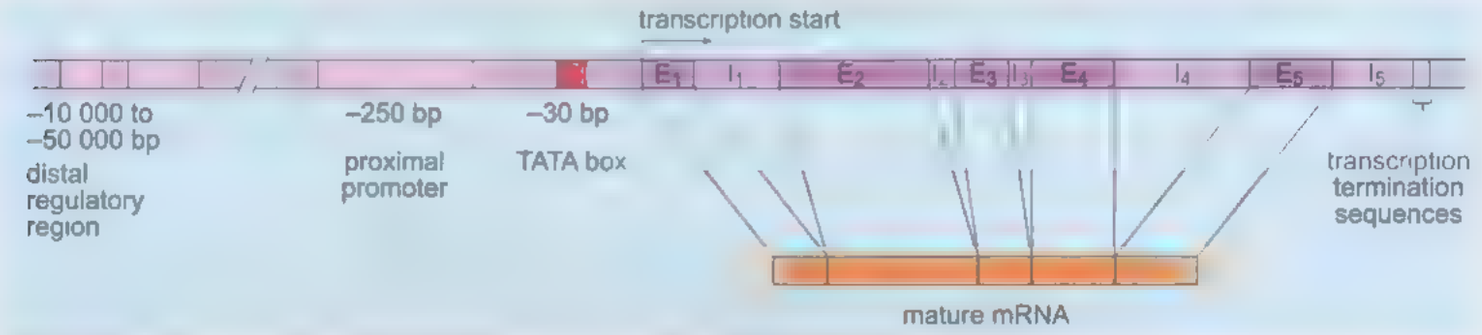



**Figure 5.26** Structure of a typical eukaryotic gene. The coding region shown has five coding exons (dark purple) with intervening non-coding regions (introns, light purple). About 30 base pairs (–30 bp) 'upstream' of the transcription start site is the TATA box (red) where RNA polymerase binds, and further upstream (–250 bp) is a proximal promoter region (pink) where protein factors that regulate transcription bind. There may also be more distal (remote) regulatory regions many thousands of bases upstream (or sometimes downstream) of the coding region to which further regulatory factors bind. Downstream of the coding region are sequences that regulate transcriptional termination (green).

Introns may be very long, and very numerous. For example, the human gene that encodes a protein called dystrophin, which is mutated in the genetic disorder Duchenne muscular dystrophy, is the longest known human gene. It is about 2.4 million ($2.4 \times 10^6$) bp in length with 79 exons, but its mature transcript after splicing is only 14 kilobases ($1.4 \times 10^4$ bp) long. The gene is so large that it is estimated that it takes 16 hours to transcribe – a particularly lengthy process compared with the duration of a typical cell division cycle (which in a cultured mammalian cell is about 20 hours). The existence of

introns permits another type of subtle regulation of gene activity: in addition to regulation at the level of transcription, several different forms of a protein may be derived from a single gene by splicing together different combinations of exons (Chapter 6).

There is a wide variety of regulatory DNA sequences both upstream and downstream of, and even sometimes within, a eukaryotic gene. These elements typically include the *promoter region*, which lies just upstream of the transcriptional start of the gene, where transcription factors assemble to promote binding of the RNA polymerase. There are also numerous more distant regulatory elements which act in concert to regulate gene expression in response to various types of signal that the cell receives from its environment (Chapter 6).

### Multiple-copy genes

Not all eukaryotic genes are present in just one copy per haploid genome: some genes, particularly those that encode cellular components required in large quantities, may be present in multiple copies. Examples include genes for the histone proteins involved in the packaging of DNA into chromatin, and the genes encoding the RNA components of the ribosomes (rRNAs), which are present in multiple copies clustered together in large arrays (Section 3.4.3). Such multiple-copy genes generally display very little sequence variation between individual gene copies, implying that mechanisms exist to prevent the accumulation of mutations in these important genes.

Other multiple-copy genes may be members of *gene families* – genes with related but often distinct roles, which have similar sequences reflecting their shared evolutionary origin. Examples of gene families and their origins are discussed later in this section.

## 5.8.2 Eukaryotic genome composition

Figure 5.27 illustrates the composition of the human genome. Notice that only a small proportion of the DNA in the human genome, about 1.5%, indicated by the dark purple segment, encodes protein or functional RNA coding sequence. From the description of eukaryotic gene structure above, you will already appreciate that the overall size of a gene can be very much larger than its protein-coding capacity would indicate, due to the presence of introns and the extensive sequences that play an important role in gene regulation. These sequences comprise about 24% of the genome and are collectively indicated by the pale purple segment in the figure.

Perhaps what is more surprising is the large proportion of the genome that is composed of repeated DNA sequences. Some of this (15.5%) comprises multiple repeats of short sequences, which are often clustered in a particular chromosomal location. The rest of the repetitive DNA (44%) is more complex and has a more variable distribution. This category largely consists of *transposable elements*, or *transposons* (described below); these are DNA sequences that, in certain circumstances, are able to excise and reintegrate elsewhere in the genome (see below). The remainder of the genome comprises unique non-coding sequences of unknown function; however, at least some of

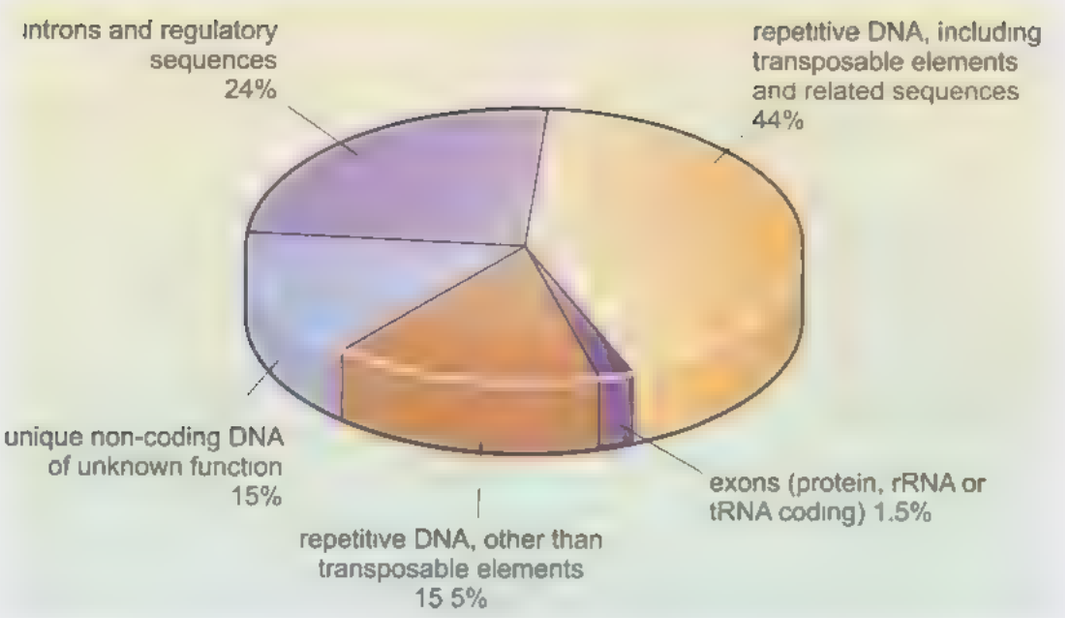



**Figure 5.27** A pie chart showing the estimated composition of the human genome (total about $3 \times 10^9$ bp). Only about 1.5% of the total DNA content actually encodes protein or functional RNA sequences.

these sequences have been conserved between species over millions of years of evolution, and some of them are transcribed, suggesting that rather than being 'junk' DNA they have as yet undiscovered functions.

## Transposable elements

So far in this book, genes have been considered as having a fixed locus (chromosomal location). However, some DNA sequences are not always found in the same position in individuals of the same species, but have the remarkable property of being mobile: that is, they can move from one locus to another within a genome. Essentially, these mobile elements can be thought of as molecular 'parasites' which often appear to have no specific function other than to maintain themselves. Such **transposable elements** (also referred to as **transposons**) are widespread, and appear to exist in all organisms, both prokaryotes and eukaryotes. They move by the process of *transposition*, a 'cut and paste' mechanism that depends on breakage and rejoining of the DNA strands.

The P element transposon, found in the fruit fly *Drosophila*, is the best studied of all transposons. The complete P element is 2.9 kilobases in length and is usually present in multiple copies, about 30–50 per cell, in so-called 'carrier strains' of *Drosophila*. The P element sequence is bounded by short DNA repeats which are essential to the transposition process, as is the single gene carried by the P element, which encodes an enzyme called a *transposase*. The transposase binds to the short repeat sequences at each end of the element and catalyses the excision and reintegration of the transposon in another location. There are many more classes of transposon for which the mechanisms of transposition vary, sometimes resulting in the multiplication of

transposon copies within a cell. Some transposons appear to be closely related to viruses.

- Suggest a likely consequence of a transposable element insertion within a gene's ORF.

  It would probably disrupt expression of the gene, causing a mutation – most likely a loss of function mutation (Section 5.6.2).

### Simple sequence repeats

A considerable proportion of a typical eukaryotic genome comprises thousands of short repeat sequences. These are often clustered around structural features of chromosomes such as the *centromere*, the region where sister chromatids attach to the spindle fibres during cell division (Activity 4.1). Such short sequence repeats are often referred to as *satellite DNA*, a name which derived from experiments in which it was found that satellite DNA could be separated from the non-repetitive DNA on the basis of its density by the technique of centrifugation. The term satellite DNA is often used interchangeably with the term heterochromatin, which describes sections of the genome that remain highly condensed during interphase, and which generally contain few genes.

## 5.8.3 Eukaryotic chromosome replication and telomeres

Earlier in this chapter, the mechanisms of DNA replication were discussed. In the case of circular DNA molecules, such as the typical prokaryotic genome, divergent replication forks originating from a single point of origin will eventually converge at a point approximately opposite the origin, leading to the complete replication of the entire circular molecule. In contrast, a typical eukaryotic genome consists of linear DNA molecules, which presents a problem in completing DNA replication. Although the leading strand can be copied right up to the end of the chromosome (Section 5.3), the situation differs for the lagging strand. There is a short stretch of DNA at the extreme 3′ end of the lagging strand, left by degradation of the final RNA primer, which cannot be filled in by DNA polymerase because there is no free 3′ OH available to initiate DNA synthesis (Figure 5.28a). Thus, during each round of DNA replication, the length of each chromosome decreases very slightly. This shortening over many cell divisions would eventually cause problems if coding regions became truncated.

The problem is overcome by the existence of a protective region of multiple short nucleotide repeats known as a **telomere**, which is present at the ends of each chromosome (Figure 5.28b). In human chromosomes, the short repeat sequence $(TTAGG)_n$ is repeated many thousands of times in the telomere.

- As a short segment at the end of each lagging strand cannot be replicated, what happens to the telomere regions during repeated cell divisions?

  The telomeres themselves gradually get shorter.

The gradual shortening of telomeres over time in dividing somatic cells has been suggested as one of many theories for why organisms age. In cell

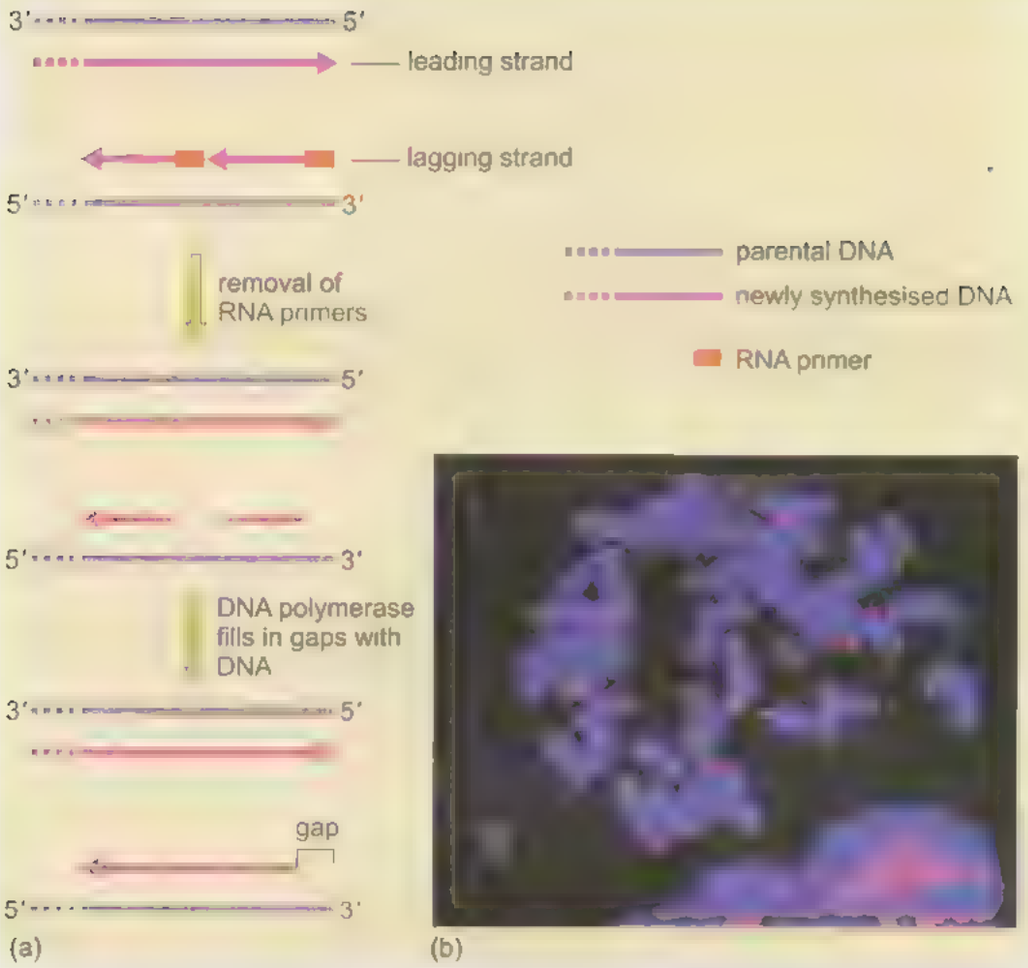



**Figure 5.28** Telomeres protect the ends of DNA. (a) Because DNA synthesis requires priming, lagging-strand synthesis omits a few nucleotides at the termini of linear eukaryotic chromosomes. (b) Human chromosomes stained to show DNA (blue) and the repetitive DNA found at telomeres (pink).

culture, eukaryotic cells only divide for a certain number of generations (30–50) before they enter a state of cellular *senescence* in which the cells remain alive but are unable to proliferate any further. This seems to be related to the gradual reduction in the length of their telomeres. In certain types of eukaryotic cells, the problem of telomere shortening is overcome by a special enzyme known as **telomerase**, which adds additional nucleotide repeats to the end of the chromosomes, thereby constantly restoring telomere length. Telomerase is very active in *stem cells*, a class of 'immortal' cells in multicellular organisms. Stem cells are able to divide continuously to provide a source of the cells that differentiate into specialised somatic cell types (Book 3, Chapter 1).

- In which other types of cells might you expect telomerase to be active?

  In germ-line cells, which in eukaryotic organisms proliferate throughout life to form the gametes, which carry the hereditary material into the next generation.

Telomerase genes also become activated in many cancer cells, which may as a result become immortal and continue to divide inappropriately, causing tumours. The majority of somatic cells, on the other hand, have no telomerase activity. While this may make them vulnerable to ageing, it may also minimise the development of cancers, by preventing uncontrolled cell proliferation.

## 5.9 Viral genomes

Viruses are, strictly speaking, not organisms since they depend absolutely on a host cell to provide most of the biochemical functions they require in order to proliferate. This is the only unifying feature of viruses: as a group, they vary widely in genome structure and size, and in the details of their propagation cycles.

While all prokaryotic and eukaryotic organisms use double-stranded DNA as their genetic material, viruses have a much wider diversity of genome structure. Viral genomes may be composed of single- or double-stranded RNA or DNA in one of several structural forms. Most viral genomes encode 'coat' proteins which enclose the genome in a protective *capsid* when they escape from the host cell to infect further cells. Some viral genomes also encode specific biochemical functions required for their propagation. For example, retroviruses make a DNA copy of their RNA genome after infecting a host cell in a process called *reverse transcription*. The DNA copy of the retroviral genome then inserts into the host cell genome where it is later transcribed to produce new viral RNA genomes. The virally encoded enzyme responsible for the synthesis of DNA from an RNA template, *reverse transcriptase*, has become an important part of the laboratory toolkit of enzymes used in manipulating nucleic acids *in vitro*.

Some viruses, such as the influenza virus, have genomes that are in several separate segments and, if an organism is infected by more than one strain of the virus, re-assortment between the genome segments can lead to the sudden appearance of new and virulent strains. This has been the case for several highly virulent epidemics of human flu viruses that contained genes from viruses that normally infect other organisms, such as swine or avian flu virus. You will read more about the interaction of viruses and host cells, and how this causes disease, in Book 3, Chapter 3.

## Summary of Sections 5.8 and 5.9

- In general, the genes and genomes of eukaryotes are larger and more complex than those of prokaryotes.
- Eukaryotic genes have a greater range of regulatory sequence elements (to which proteins bind and regulate transcription), and their ORFs usually consist of several short coding sections (exons) interrupted by non-coding regions (introns) which must be removed by RNA splicing to produce a mature mRNA for translation.

- Eukaryotic genomes are composed of multiple linear chromosomes, which have multiple replication origins and have special structures, telomeres, at their termini to protect them against gene loss due to incomplete replication of the lagging strand.
- The genomes of eukaryotes contain large amounts of DNA with little apparent function, including repetitive sequences and transposable elements (transposons).
- Viruses are not regarded as organisms, and depend on the host cell for most of the biochemical functions they require to proliferate. Viruses have a wide diversity of genome structures including RNA as well as DNA genomes.

## 5.10 [illegible] Genomics

For many years, researchers were only able to estimate the numbers of genes in more complex genomes by extrapolating from the simpler, better characterised systems, which led to estimates that erred very much on the high side. Advances in laboratory techniques for DNA cloning and DNA sequencing have since enabled scientists to determine the complete genome sequences of a growing number of species. As more genomes are sequenced, they add to a rich set of data that increasingly informs the biological understanding of evolutionary relationships between taxonomic groups, based on similarities between gene sequences and the patterns of genome rearrangements that have occurred through time.

### 5.10.1 A history of genome sequencing

The early efforts to sequence genomes were limited by the technology of the time to small genomes, such as bacteriophage and mitochondria. By the 1990s, instruments that automated the collection of sequence data had significantly speeded up the process. The first organism to have its genome fully sequenced was the bacterium *Haemophilus influenzae* in 1995, and the first eukaryote was *Saccharomyces cerevisiae* (baker's yeast) in 1996. Sequencing of the human genome began in 1990; a draft sequence was available by 2000, and the 'complete' sequence by 2003. Continued technical development has significantly increased the speed and reduced the expense of genome sequencing (Box 5.3).

#### Box 5.3 DNA sequencing

The technique of DNA sequencing is based on the principle that DNA fragments of different length can be separated from one another by a technique known as gel electrophoresis. A thin layer of gel containing ionic salts (which allows the gel to conduct electricity) is formed between two plates. The mixture of different-length DNA molecules to be separated is layered into a 'well' at one end of the gel and an electric current is applied through the gel, causing the negatively charged DNA molecules (negative because of their negatively charged phosphate groups) to migrate through the gel towards the positive electrode

Smaller fragments move more easily through the gel matrix (which acts rather like a sieve), so they move further away from the origin towards the other end of the gel, while larger DNA fragments move less far. By choosing the gel composition appropriately, gel electrophoresis can easily discriminate between fragments that differ in length by as little as a single nucleotide.

The detailed method and technical implementation of DNA sequencing is complicated, and subject to continual advances in instrumentation. For the purposes of this chapter, the details are not as important as the principle. In order to appreciate how the technique works, you need to recall some of the features of DNA structure and DNA replication. Earlier in this chapter, you learnt how DNA is replicated by successive linkage of nucleotides to a growing strand via complementary base pairing to a single-stranded template DNA molecule. Crucially, the synthesis of a new DNA chain can be terminated if a defective nucleotide is incorporated into the growing DNA chain.

The principle of the most commonly used method called *chain termination sequencing* is to carry out DNA replication *in vitro*, but to 'spike' the reaction mix with a small amount of 'defective' forms of dNTPs called ddNTPs (dideoxynucleoside triphosphates) or more specifically ddATP, ddTTP, ddGTP or ddCTP. The ddNTPs are DNA chain terminators – when they are incorporated into the DNA chain no further nucleotides can be added because they lack a 3′ –OH group required for the formation of a covalent phosphodiester bond with another nucleotide.

For chain termination sequencing, four *in vitro* replication reaction mixtures are prepared, each containing the template DNA, a short DNA primer, a mixture of the four dNTPs, DNA polymerase and a small amount of one of the ddNTPs, e.g. dideoxyadenosine triphosphate (ddATP). Whenever the elongating DNA chain incorporates a ddATP (instead of a dATP), synthesis terminates. By ensuring that the concentration of ddATP in the reaction is very low relative to the concentration of dTTP, dATP, dCTP and dGTP, a range of DNA fragment lengths will be synthesised, all of them terminating with the incorporation of a ddATP. In other words, each of the newly synthesised fragments will terminate at one of the points where there is a T in the DNA template sequence (Figure 5.29a). Three similar reactions are set up containing ddTTP, ddCTP, or ddGTP, which will thus provide a range of fragments terminating where A, G and C, respectively, occur in the template sequence. The reactions terminating with incorporation of ddATP, ddGTP, ddTTP and ddCTP can be referred to as the A, G, T and C reactions, respectively (Figure 5.29b).

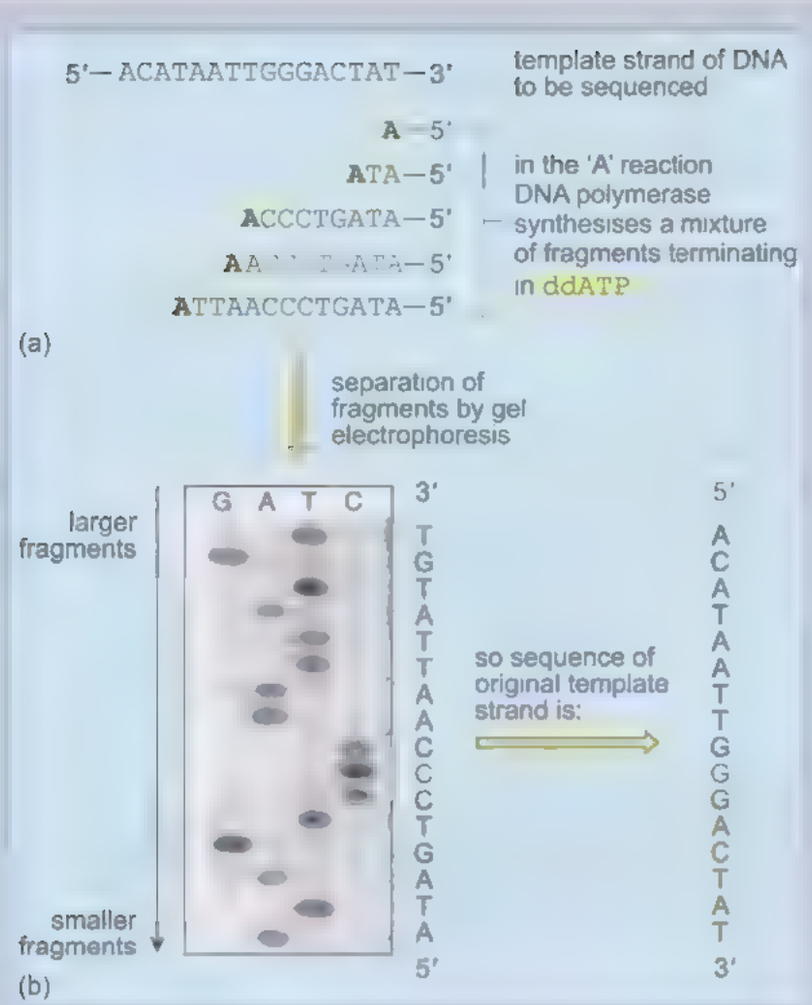



**Figure 5.29** Chain termination sequencing. (a) The principle by which the 'A' reaction generates a mixture of fragments ending in ddATP, representing each place in the template DNA where there was a T. (b) Four DNA synthesis reactions containing a mixture of DNA fragments terminating in ddGTP, ddATP, ddTTP or ddCTP are separated by gel electrophoresis in separate lanes as indicated. Smaller fragments move further down the gel. The sequence of the synthesised strand can thus be read from the bottom to the top of the gel.

In order to 'label' the fragments and make them visible after gel electrophoresis, a small amount of dNTPs labelled with radioactivity can also be added to each of the four reactions to be incorporated into the newly synthesised DNA fragments.

The reaction products are then separated by gel electrophoresis, as shown in Figure 5.29b and the separation pattern of the newly synthesised DNA fragments is revealed by exposing the gel to X-ray film, which darkens where it is exposed to radioactivity. The DNA sequence can be 'read' from the order of the bands on the gel.

DNA sequencing technology has continually improved to enable higher throughput. Nowadays, automated sequencing uses ddNTP terminators labelled with different coloured fluorophores, so that all four reactions

can be carried out in a single tube. This innovation also means that sequence data can be read automatically in the gel as the fragments migrate past a fluorescence detector attached to a computer, allowing complete automation of the procedure

The chain termination method of DNA sequencing is limited by the size-resolving capacity of gels and can elucidate only a few hundred base pairs of sequence in each gel run. This is very small in comparison to the size of even a prokaryotic genome, so a great many sequencing gels are required to determine a complete genome. Most genome sequencing projects take a 'shotgun' approach, breaking up genomic DNA into random small fragments, which are used to determine a large collection of short DNA sequences, each of which overlaps with some other sequences. The sequences are then analysed, sorted and reconstructed by computer, using software that compares each short segment of sequence against all the others, and generates larger contiguous sequences based on overlapping short sequences. Typically, each base pair of a finished genome sequence will have been sequenced many times, being represented in numerous overlapping sequences. Once the genome of a particular species has been sequenced, this sequence can be used as a 'scaffold' to much more quickly assemble genome sequences of other individuals of the same species. This has been an important factor in developing high-speed sequencing of human genomes.

The rapid advances in sequencing technology have made genome sequencing both quicker and less expensive. Not long before the start of the Human Genome Project in 1990, typical sequencing projects cost approximately $1 per base pair of 'finished' genome sequence (thus a genome of $3 \times 10^9$ bp would have cost around 3 billion US dollars). At the time of writing (early 2012), an individual's DNA sequence could be determined for only a few thousand dollars, and the future likelihood of affordable routine genome sequencing for everyone may herald a new era of 'personalised medicine' in which genetics could be used to predict susceptibility to disease and response to treatments. The ethical debate surrounding the ability to predict an individual's heritable characteristics is not something there is space to explore here, but it is likely to have profound effects on the use of such technology.

From a genetic perspective, one of the major outcomes of genome sequencing, particularly of the human genome, has been the observation of widespread genetic variation not only between species, but also between individuals of the same species.

## 5.11 Variation and evolution of genomes

As the genomic DNA sequences of an increasing number of species have been determined, comparisons of gene sequences within and between genomes has helped to reveal the functions and origins of many genes and has provided some of the strongest evidence supporting phylogenetic relationships during evolution (Section 1.2.2). Genes in different species that share a degree of sequence similarity due to their common evolutionary origins are known as

**homologous genes** (the term 'homologous' is often used more loosely, and incorrectly, to mean 'similar'). If the DNA sequence of a specific human gene is compared to its mouse counterpart, although the two genes clearly encode very similar proteins, they are not identical. That is, there are DNA sequence differences between them. How do genomes change in sequence during the evolutionary history of a species?

DNA replication and repair mechanisms are remarkably accurate but errors occur, albeit at a low rate in most organisms.

- What is a typical overall error rate during mammalian DNA replication?

- Typically one error per $10^9$ bases copied (Section 5.3).

The 1000 Genomes Project (launched in 2008) is an international project which set out to determine the DNA sequence of at least 1000 human genomes (and indeed set its sights at a higher final total), and thereby identify the genome locations where there is genetic variation in human populations. Much of the genetic variation within *H. sapiens* takes the form of **single nucleotide polymorphisms (SNPs)** – sites in the human genome where there are single nucleotide differences between individuals.

- What are some of the processes that can result in variation at a particular locus in the genome among individuals.

  Variation can result from misincorporation of bases during replication, nucleotide insertions or deletions and base changes due to chemical modification (e.g. deamination, Section 5.6.1).

The majority of SNPs, though by no means all, are single nucleotide changes which have little or no effect on gene activity.

The second major type of variation is gene **copy number variation (CNV)**. Some genes may be present in a variable number of copies in different individuals of the same species, resulting in different levels of expression of the gene product. CNV is a result of the duplication of existing genes, or larger segments of the genome, over time. Gene duplication can also result in the evolution of *gene families*, structurally related genes which may have similar or different functions.

### 5.1.1 [illegible]

There are a number of mechanisms by which segments of DNA may be duplicated, including errors during DNA replication, unequal crossing over between two homologous chromosomes during meiosis (Section 4.3.1 and Activity 4.1), or the action of transposons (Section 5.8.2). Transposons may carry a section of flanking genomic DNA with them when they excise from the genome and insert in another location, in some cases effectively duplicating some of the host's genes. Over long periods of time, a single ancestral gene may therefore undergo a number of different duplication events. Occasional duplication of a whole chromosome or indeed of the whole genome (see below) brings about the duplication of many genes simultaneously.

Once a gene encoding a particular product is duplicated, redundancy is created. In other words, as long as one copy of the gene continues to operate normally and produce its essential gene product, the other gene copy is surplus to requirements and is freed from selection pressure. One of the newly duplicated genes may therefore change in a way that prevents it from producing its original product without any deleterious consequences for the fitness of the organism. Mutation and natural selection can thus result in divergence of the sequence of the two genes during subsequent generations, each gene can follow its own evolutionary path. In many cases, one of the duplicated genes simply acquires one or more mutations that prevent production of, or abolish the function of, its product. Such a permanently inactivated gene is known as a *pseudogene*. One of the duplicates may, however, acquire a new function. A **gene family** is a collection of structurally related genes (some may be functional, some may be pseudogenes) that have appeared during evolution by means of repeated duplication and divergence from a single ancestral sequence.

An example of a gene family is that of the human globin genes. Globins are proteins that play an important role in oxygen transport in all vertebrates, and in many invertebrates. The simplest form of globin, found in certain marine worms, a few insects and some fish, is a monomeric (single polypeptide) protein encoded by a single gene. Human haemoglobin, the major component of red blood cells, is in contrast a tetrameric (four polypeptide) unit composed of two α (alpha) globin and two β (beta) globin polypeptides (Figure 5.30). Each type of α and β chain is encoded by a corresponding α and β gene that arose by duplication followed by mutation of a single ancestral globin gene. During evolution, both α and β genes have in fact undergone several distinct duplication events. As a result, in humans there are a number of α genes arranged in a cluster on chromosome 16, while a number of β genes are clustered separately on chromosome 11. Both groups include non-functional pseudogenes, which are denoted by ψ (the Greek letter 'psi') within their names. The closely related gene myoglobin, lies on chromosome 22. Myoglobin has similar structure and properties to haemoglobin, but is found in muscles and is monomeric.

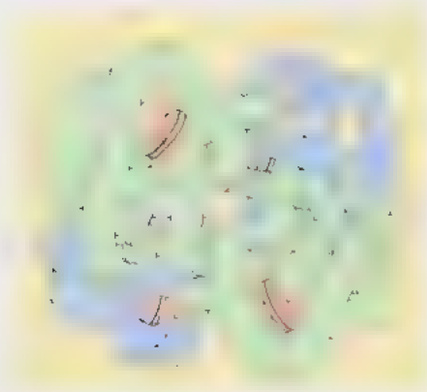


**Figure 5.30** Schematic representation of the four-chain vertebrate haemoglobin (with two α and two β chains). Each of the red disks represents a haem group – the iron-containing component of haemoglobin, which carries oxygen.

Figure 5.31 shows the evolutionary history of the oxygen-carrying globin genes in humans, deduced from sequence similarity in different species, starting with the single chain globin gene at the top and finishing at the bottom with the α- and β-globin gene clusters and myoglobin found in current-day primates. The approximate dates at which each gene duplication event is thought to have occurred are shown at each branch point in the diagram, and are correlated with major events in animal evolution. First, the single ancestral gene was duplicated. One copy underwent no further duplication, and became the myoglobin gene. The other copy of the globin gene underwent further duplications later in evolution, yielding the genes of the α- and β-globin families, which were favoured by natural selection and so were retained.

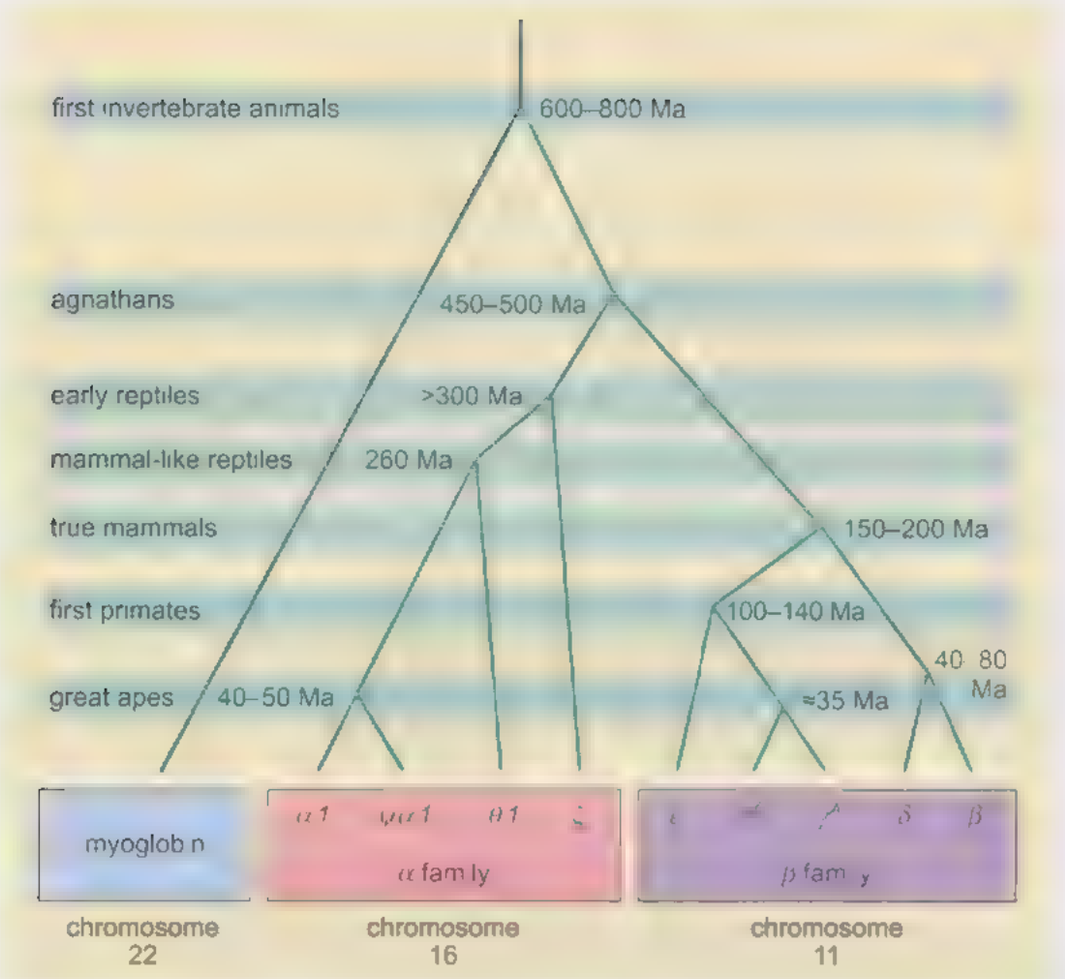



**Figure 5.31** The ancestry of the human globin genes deduced from sequence comparisons. The timescale on the diagram runs from top to bottom, i.e. more ancient gene duplication events lie towards the top, and more recent events towards the bottom. The appearances in the fossil record of new animal taxa are indicated as horizontal blue bars, but these do not correspond specifically with the globin gene duplication/divergence events.

## 5.11.2 Chromosome and genome duplications

Duplication events are not confined to single genes. When chromosomes of closely related species are compared, large chromosome rearrangements are frequently observed. Chromosome rearrangements include duplications (where a section of chromosome has been duplicated), deletions (where a section of chromosome has been lost) and inversions (where a section of chromosome has been flipped in orientation relative to the 'normal' chromosome order).

The duplication of sections of genomes containing many genes (known *segmental duplication*) is quite a common event. Segmental duplications may be located on the same chromosome or on different chromosomes. In some cases, the segments of duplicated DNA are very large, sometimes several hundreds of thousands of base pairs in length. One of the more surprising findings from human genome sequencing projects is that almost 5% of the human genome is made up of this type of duplication. Individuals differ widely in the detail and extent of individual duplications: in some cases, many

copies of a segment are present. We currently know very little about how or why these segmental duplications have arisen, but they have probably played a significant role in the creation of new human genes and gene families during evolution.

Whole genome duplication and even triplication are relatively frequent events in the evolution of plant species. Notably, modern wheat varieties are known to be hexaploid derivatives of their ancestral forms; that is, the genome has triplicated from the initial diploid state to give six copies of the haploid genome. While most common in plants, cases of apparent genome duplication in animals are known. For example, a whole genome duplication appears to have occurred during the evolution of bony fish, and more recently in the genome of the African clawed frog *Xenopus laevis*, which is double the size of that of other living species in the genus, implying that a genome doubling event occurred at some time after the appearance of the genus *Xenopus*.

### 5.11.3 Genome-wide association studies (GWAS)

Conventional genetics experiments are impossible to perform on humans, for obvious ethical and practical reasons, but the identification of large numbers of SNPs, CNVs and other genetic variants as a result of genome sequencing has brought with it a new era of human genetics.

Many millions of SNPs have been identified by comparing individual human genome sequences. Genome-wide association studies (GWAS) make use of SNPs to look for an association between genetic variation and certain traits, such as the risk of developing a disease. SNPs only very occasionally impact directly on gene expression, the great majority of SNPS have no discernible phenotypic impact. However, while the majority of SNPs are not directly associated with causing a disease or susceptibility to a disease, those that lie very close to (i.e. are linked to) disease-related genes can be used as 'markers' in genetic linkage studies to identify whether a disease-related allele is present. SNPs can thus be thought of as silent mutations that provide a way of 'mapping' an individual's SNP genotype, i.e. the particular combination of SNPs in their genome, and whether it correlates with a particular characteristic, such as disease susceptibility.

Recall from Section 4.4.3 that the recombination frequency between two loci is related to the physical distance between them. Therefore genes and nearby SNPs that are very close together on the chromosome will tend not to be separated by recombination. A disease-related gene allele can therefore often be followed through the generations just by detecting a particular combination of SNP variants that remain linked to it. Such a combination of SNPs is known as a *haplotype*. Just identifying a few SNP variants in a stretch of several hundred thousand bases can be enough to confirm that a particular gene allele is present, which is much more efficient than sequencing the whole DNA region in every individual. Typically, a GWAS uses high-throughput analytical techniques to allow researchers to sample several hundred thousand known SNP locations in each individual subject, and then rapidly analyse the pattern of variation across the whole genome. These studies can identify the genetic contribution to disease by comparing the DNA of two groups of

participants: people with the disease and a similar group of people without the disease. The SNP genotype of the DNA from each individual is determined and analysed to reveal any association between a particular SNP haplotype and the group that have the disease (Figure 5.32).

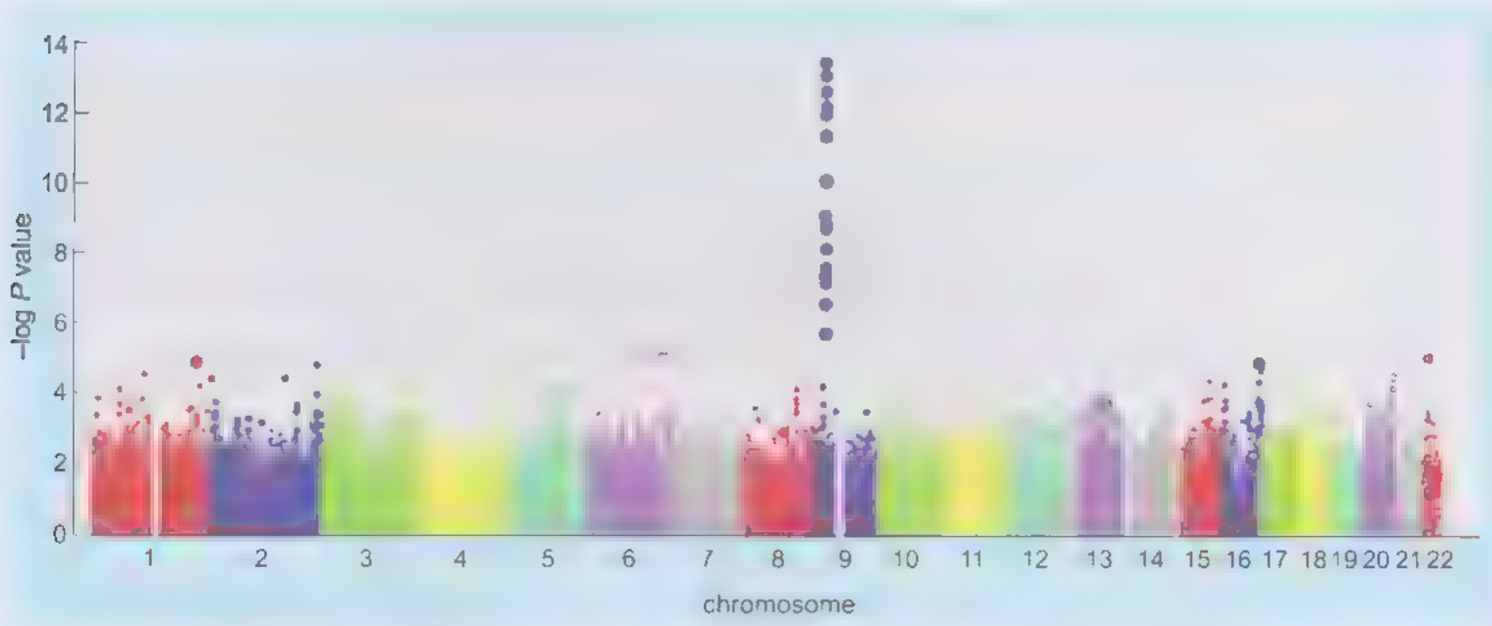



**Figure 5.32** A genome-wide association study of single nucleotide polymorphisms associated with myocardial infarction (commonly known as a heart attack). This type of plot is referred to as a Manhattan plot. The *x*-axis refers to several thousand locations along the human genome (by chromosome, 1 to 22) where there are known to be SNPs, each represented by a dot. The *y*-axis refers to probability (the *P* value) that the presence of a particular SNP is associated with the presence of the disease. You can see that variation in a particular SNP located on chromosome 9 is associated with a high probability of the disease.

To date, these studies have identified both genetic risk and protective genetic factors for asthma, cancer, diabetes, heart disease, mental illness, and several other human diseases. GWAS therefore offers the opportunity to take a combined molecular and genetic approach to identifying the gene variants that contribute to many medical conditions.

## Summary of Sections 5.10 and 5.11

- Comparison of the genomic DNA of different species has helped to reveal the function and evolutionary origin of many genes, and also the enormous genetic variation between individuals of a species.
- The two main types of genomic variation are single nucleotide polymorphisms (SNPs) and gene copy number variation (CNV).
- During evolution, gene copy numbers may change via duplication of genomes, segments of genomes or individual genes. This is how gene families have arisen.
- Genome-wide association studies (GWAS) can help to identify linkage between genome variation and complex traits such as disease susceptibility.

## 5.12 Final word

This chapter has described the structure and function of prokaryotic and eukaryotic genomes, and how the genetic information within the genome is accurately copied and passed on to daughter cells. The genomes of prokaryotic organisms are relatively compact and small, enabling these organisms to respond rapidly and efficiently to environmental challenges. In contrast, most eukaryotic genomes are larger and more complex. Despite these differences, eukaryotes and prokaryotes share many systems by which replication errors and damage to DNA may be corrected or repaired. Nevertheless, occasional sequence changes are inherited and give rise to mutations. Advances in genome analysis techniques, summarised in this chapter, have begun to unravel the properties and evolutionary history of complex eukaryotic genomes.

The next chapter looks in more detail at the control of gene expression, and the essential role played by some non-coding DNA sequences.

## 5.13 Learning outcomes

5.1 Explain how the structure of DNA relates to the mechanisms of DNA replication.

5.2 Describe the roles of the major components of the replication complex in DNA replication.

5.3 Describe the mechanisms by which the accuracy of DNA replication is ensured, and by which damaged DNA can be repaired.

5.4 Use the genetic code to determine the sequence of a gene product.

5.5 Explain the effects of different classes of mutation on the expression or function of a gene product.

5.6 Compare the general features of gene structure and genome composition in prokaryotes and eukaryotes.

5.7 Outline some of the techniques used to study gene sequence and function.

# Chapter 6 The control of gene expression

## 6.1 Introduction

This chapter considers gene expression, the process by which the information in the genome of an organism is used to synthesise gene products, which are either proteins or functional RNAs. The gene encodes the linear sequence of a gene product, but in addition, the DNA flanking the coding region contains much more information in the form of regulatory DNA sequences that specify when, and in what type of cell, the gene product is expressed. Precise regulation of gene expression is crucial. If a protein is produced in a cell at the wrong level or at the wrong time, the effect may be devastating.

You should bear in mind an important difference between the role of gene regulation in single-celled prokaryotes and in multicellular eukaryotes. Gene regulation allows bacterial cells to adjust rapidly to changes in their environment, and to optimise their growth and proliferation. Many eukaryotic genes are also regulated in response to external environmental changes, but additionally gene regulation in multicellular organisms drives the development of the different types of specialised cells. All of the cells in a multicellular organism originate from a single cell (the zygote) and therefore all carry the same genetic information, but during development of the organism different patterns of gene expression result in different types of cell with unique properties.

- What is the alternative name for this process of cell specialisation?

  Differentiation.

So, how is gene expression controlled? In order to answer this question, it is first necessary to examine in more detail the processes by which the gene products (i.e. RNAs and proteins) are synthesised using the information encoded in DNA. The first part of this chapter will recap the processes by which proteins are synthesised, and the remainder of the chapter will look in more detail at the mechanisms of gene expression and how it is regulated in prokaryotes and eukaryotes.

## 6.2 An overview of gene expression

You will recall from Chapter 4 that a gene is a heritable unit that contributes to the physical characteristics (the phenotype) of an organism. It consists of a segment of genomic DNA that specifies the structure of a polypeptide or a functional RNA (e.g. a ribosomal RNA or a transfer RNA). In Activity 5.1 Parts 1 and 2 you observed how complementary base pairing involving hydrogen bonding between nucleotide bases in the two strands of the DNA double helix plays an essential role in: (1) the stability of the DNA molecule, (2) accurate replication to produce two identical DNA double helices (Section 5.3), and (3) the process of gene transcription.

Figure 6.1 shows the two main stages in the information flow from DNA to protein. In the first step, transcription, a segment of sequence on only one strand of genomic DNA (the template strand) is transcribed to produce an intermediary RNA molecule called a messenger RNA (mRNA). Complementary base pairing between the DNA template strand and ribonucleotides produces an mRNA that is complementary to the DNA template strand. In the second step, translation, the sequence of the mRNA is translated into the amino acid code, to produce a polypeptide.

- What unit along the mRNA strand encodes a single amino acid in the polypeptide chain?
- Each 'triplet' of three bases in the mRNA sequence, known as a codon, specifies one amino acid in the polypeptide chain.

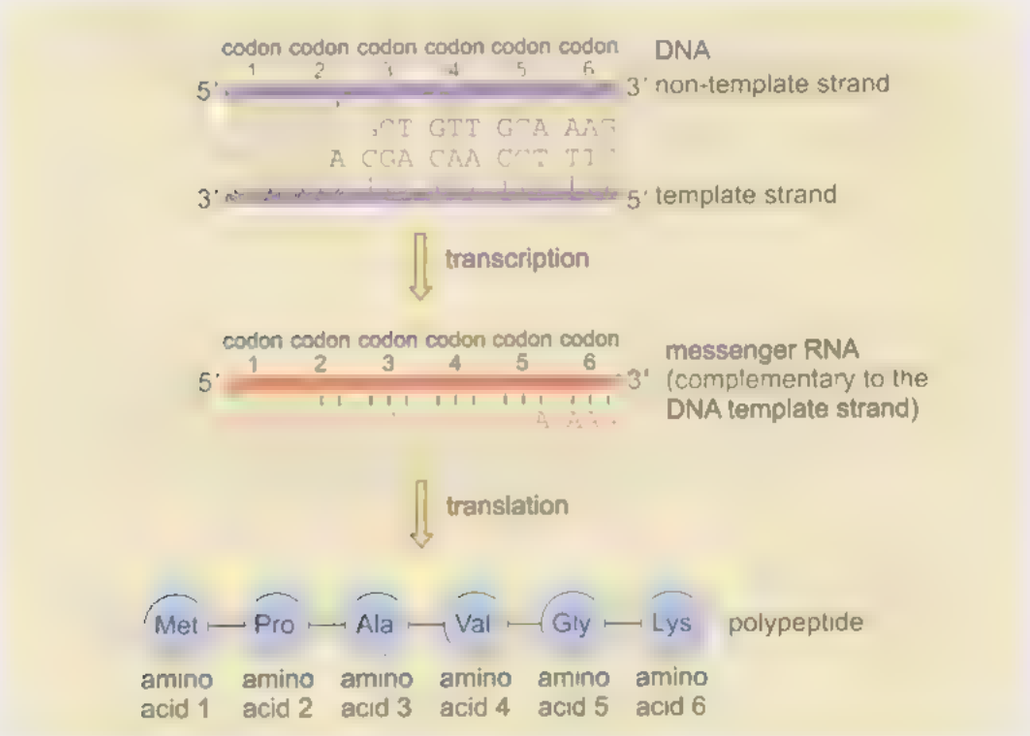



**Figure 6.1** The relationship between the codons in DNA and mRNA and the amino acids in the encoded polypeptide. The sequence of the DNA template strand is transcribed into mRNA (note that RNA has the base uracil (U) instead of thymine (T)) and the mRNA sequence is translated into the polypeptide sequence. Each triplet of bases (a codon) specifies a particular amino acid.

Note that the flow of information occurs in only one direction: from DNA sequence, via RNA, to protein sequence, and never backwards from protein to nucleic acid (DNA or RNA). This one way flow of information: DNA makes RNA makes protein, is often called the *central dogma* of molecular biology, first proposed by Francis Crick in 1958.

## 6.2 DNA into RNA

Like DNA, RNA is a linear polymer composed of four different types of nucleotides linked together by covalent phosphodiester bonds (Figure 6.2a and Section 5.2.1). However, while DNA is double-stranded, RNA is usually single-stranded (although it is able to form *internal* base pairs between bases in different parts of the single strand). The nucleotides in RNA are ribonucleotides containing the 5-carbon sugar ribose (DNA contains deoxyribose). RNA also differs from DNA in that it contains the base uracil (U) instead of thymine (T); however, since U also base-pairs with A, a DNA strand can act as a template for RNA strand synthesis by complementary base pairing.

Whilst the deoxyribonucleotides composing DNA contain the 5-carbon deoxyribose sugar (Figure 5.2a), ribonucleotides contain ribose sugar which has a second hydroxyl group on the 2′ carbon, but is otherwise identical to deoxyribose.

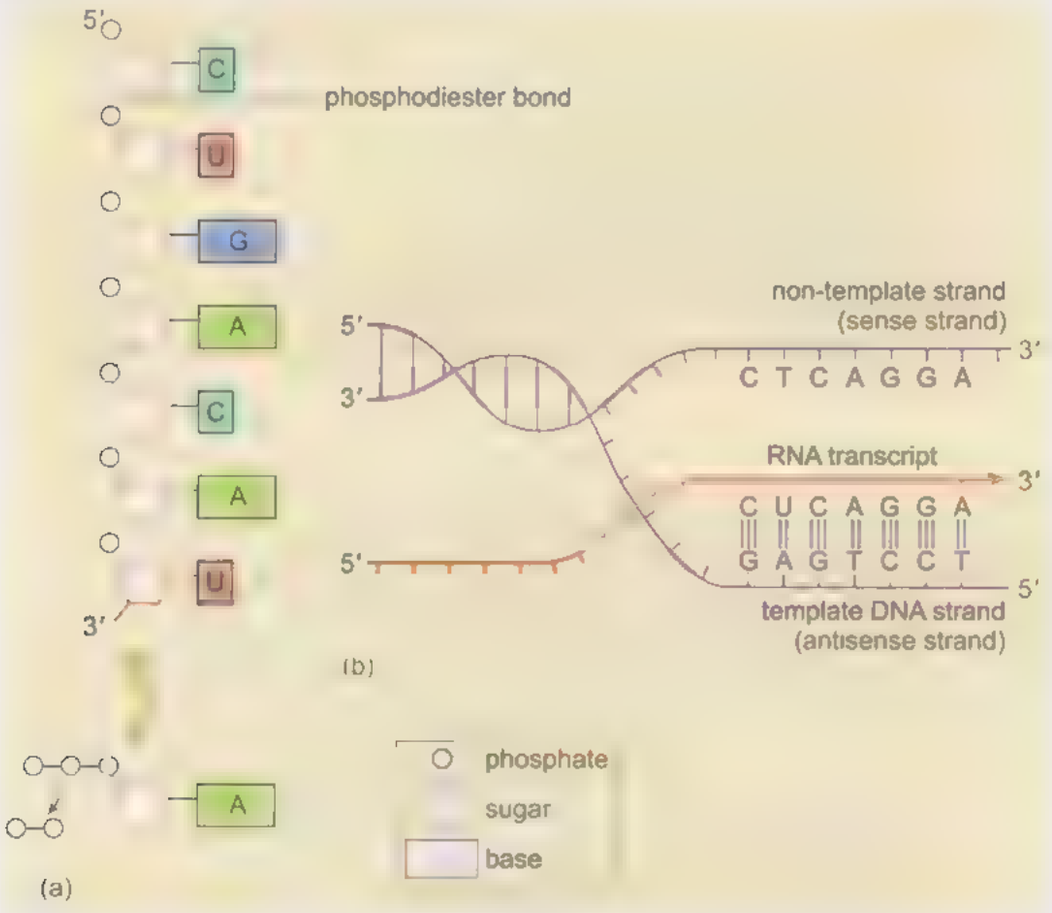



**Figure 6.2** DNA and RNA strands in transcription. (a) RNA is synthesised from ribonucleotide precursors with three linked phosphate groups. As each ribonucleotide is added to the 3′ end of the growing RNA strand, two phosphates are removed from the ribonucleotide, releasing energy for phosphodiester bond formation. (b) It is necessary to discriminate between the two DNA strands because only one acts as the template for transcription of a gene. Except for the substitution of U for T, the RNA transcript has the same sequence of bases as the complementary DNA strand (which is therefore sometimes called the sense strand, while the template DNA strand is known as the antisense strand). RNA is always synthesised in a 5′ to 3′ direction, reading the template DNA strand in the 3′ to 5′ direction.

During transcription, the two DNA strands unwind in the region that is being transcribed (Figure 6.2b). Transcription is different from DNA replication because only one of the DNA strands, the *template strand*, is 'read' to produce the single-stranded RNA copy. The other DNA strand is known as the *non-template strand* (or the sense strand, because its sequence corresponds to that of the transcribed RNA while the template DNA strand is also known as the antisense strand, Figure 6.2b).

The synthesis of RNA molecules in cells is carried out by enzymes called *DNA-dependent RNA polymerases*. Throughout this chapter, the shortened name **RNA polymerase** will be used for these enzymes. RNA polymerase binds to the DNA at the start of the gene and separates the two strands by breaking the hydrogen bonds between complementary nucleotides. Unlike DNA polymerase, RNA polymerase is capable of *initiating* polymerisation of a new RNA chain; it doesn't require a primer to provide a 3′ –OH. RNA is synthesised from precursors known as ribonucleoside triphosphates (ATP, GTP, CTP and UTP). The polymerase moves along, 'reading' the base sequence of the template DNA strand in a 3′ to 5′ direction, sequentially linking ribonucleotides to the 3′ end of the new RNA chain. Synthesis of the RNA transcript therefore always proceeds in a 5′ to 3′ direction (Figure 6.2b).

- What is the sequence of the RNA transcribed from the following template DNA strand 3′ AAGCTCGACTTGACT 5′?

  The sequence of the RNA transcript is complementary to the template strand, i.e. 5′ UUCGAGCUGAACUGA 3′.

As the RNA polymerase moves along the template strand, the two DNA strands behind it rewind into the double helix. How the RNA polymerase recognises where to start and where to stop transcribing is addressed later in the chapter.

### *The different forms of transcribed RNA*

Transcribed RNAs can be divided into two major groups, *coding RNAs* and *non-coding RNAs*. **Coding RNAs** are the transcripts of protein-coding genes.

- What is the alternative name given to a mature protein-coding RNA?

  Messenger RNA (mRNA).

Thousands of different genes are being transcribed in a cell at any one time. mRNAs are usually very short-lived molecules; the half-life of bacterial mRNA after synthesis is no more than a few minutes, and in most eukaryotic cells the majority of mRNAs are degraded within a few hours.

*Half-life* is the period of time it takes for a substance undergoing decay to decrease in amount by one-half.

- What is the advantage to a cell of producing short-lived mRNAs?

  This allows cells to rapidly adjust the rate of protein synthesis in response to their changing needs. If all mRNAs were very long-lived, they might continue to be translated when their protein products were no longer required.

mRNA, however, constitutes only about 4% of the total quantity of RNA in a eukaryotic cell. In fact, most of the RNA is **non-coding RNA (ncRNA)** – the

product of non-protein-coding genes. These ncRNAs have a range of different functions, all of which are performed directly by the RNA molecules themselves. The most abundant type of ncRNA is **ribosomal RNA (rRNA)**, which makes up over 90% of the total RNA in actively dividing cells. rRNAs are important structural components of the ribosome. Another type of ncRNA is **transfer RNA (tRNA)** which carries individual amino acids to the ribosome, where they are incorporated into a newly translated polypeptide chain (Activity 5.1, Part 2). Both of these types of RNA are produced by transcription in the same way as mRNA, but they are not translated into protein.

Eukaryotes also contain a large variety of other much shorter ncRNAs, some of which (for example, the class known as *microRNAs*) have a role in controlling gene expression (Section 6.5.4).

## mRNA into protein

Ribosomes in the cell cytoplasm bind to the mRNA and use the sequence of nucleotide bases in the mRNA sequence to specify the order of amino acids in the polypeptide chain (Activity 5.1, Part 2). As you learnt in Chapter 5, the relationship between the three-base codons in RNA and the amino acids they specify is known as the genetic code (Figure 5.13). There are 64 possible triplet codons, but only 20 amino acids occur naturally in proteins, so the excess of codon combinations means that there are actually several codons for most of the amino acids. In addition there is a 'start' codon (AUG), which also encodes methionine, and there are three 'stop' codons (UAA, UAG and UGA) that signal the termination of translation.

- Using the genetic code table in Figure 5.13, what is the sequence of the polypeptide translated from the following mRNA strand:
  5′ AUGGCCGGACAUGCUUCGCGG 3′?

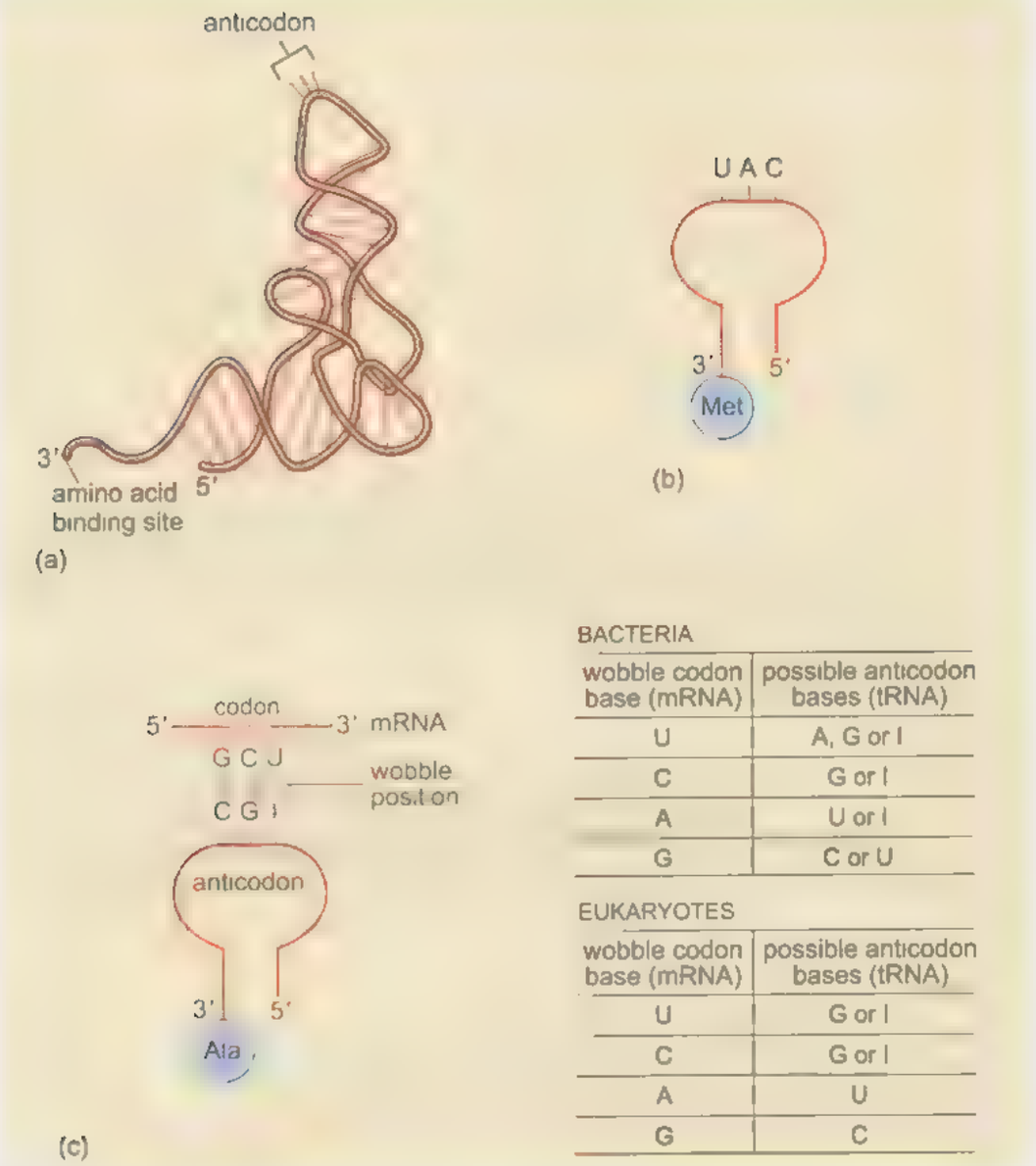
  The polypeptide is: Met Ala Gly His Ala Ser Arg (or M A G H A S R)

### *Recognition of mRNA codons by tRNA*

The link that enables conversion of the four-letter code of DNA and mRNA into the 20-letter code of proteins is transfer RNA (tRNA). tRNAs are single-stranded RNA molecules that fold up in a particular conformation (three-dimensional shape) due to internal complementary base pairing between nucleotides in different parts of the single tRNA strand (Figure 6.3a). The conformation of the tRNA is such that a sequence at the 3′ end of the tRNA is exposed, and it is here that the amino acid attaches. At another site on the tRNA is a region where there are three exposed, *unpaired* bases; these constitute the **anticodon**.

- How is the tRNA involved in translation of an mRNA sequence?

  The *anticodon* of the tRNA can interact with a complementary *codon* in mRNA by complementary base pairing (Activity 5.1). The amino acid attached to the tRNA is thus brought into position to be incorporated into a polypeptide chain.




BACTERIA

| wobble codon base (mRNA) | possible anticodon bases (tRNA) |
|---|---|
| U | A, G or I |
| C | G or I |
| A | U or I |
| G | C or U |

EUKARYOTES

| wobble codon base (mRNA) | possible anticodon bases (tRNA) |
|---|---|
| U | G or I |
| C | G or I |
| A | U |
| G | C |

**Figure 6.3** Transfer RNA (tRNA) (a) The generalised structure of a tRNA molecule showing the three-dimensional shape conferred by complementary base pairing within the single strand (b) A schematic representation of a tRNA carrying methionine. (c) The wobble base pairs that commonly form between the third (3′) base in an mRNA codon and the 5′ base of the tRNA anticodon in bacteria and eukaryotes. A 'non-Watson Crick' pairing between uracil (U) and inosine (I) is represented here as a single hydrogen bond (rather than the two or three hydrogen bonds of Watson–Crick base pairs).

A tRNA with a particular anticodon sequence always carries the same amino acid. For example, 5′ AUG 3′ is the translation 'start' codon, but it is also the codon for the amino acid methionine (Met), and is recognised by a tRNA with the anticodon 5′ CAU 3′ (Figure 6.3b). Some amino acids are encoded by several different codons. In fact, only methionine (AUG, which is also the start codon) and tryptophan (UGG) each have a single type of codon.

- Look at the genetic code in Figure 5.13 and identify all the codons for valine (Val). How do the codons compare?

- There are four codons for valine: GUU, GUC, GUA and GUG; the first two bases are identical for all of them, while the third varies.

Having several codons for some amino acids is referred to as *degeneracy* or *redundancy* of the genetic code. Degeneracy of the genetic code does not imply any ambiguity, each type of tRNA molecule can be attached to only one type of amino acid. The four codons that specify valine (Val), for example, never carry any other type of amino acid.

If there were a one-to-one correspondence between tRNA molecules and codons, all cells would require 64 different types of tRNA molecules (61 carrying amino acids and three that recognise stop codons). However, most organisms actually have only around 40–50 different types of tRNAs, because some anticodons can pair with more than one codon due to a phenomenon known as *wobble base pairing*. This occurs because the 5′ base of the anticodon (which base-pairs with the 3′ base of the mRNA codon) does not have such strict base pairing requirements as the other two bases, and can in some cases form 'non-Watson–Crick' base pairs (i.e. something other than A:U and C:G base pairs). Many tRNAs contain unusual or modified forms of bases that are particularly prone to wobble. For example, if the base at the 5′ wobble position of a tRNA anticodon is inosine (I), the tRNA can recognise any one of three different codons in bacteria and either of two codons in eukaryotes (Figure 6.3c). Hence fewer than 64 types of tRNA anticodon are actually required to recognise all of the possible codons.

Translation is *initiated* when a ribosome assembles on the mRNA at the start of a protein-coding sequence. The first tRNA to bind to the mRNA at the ribosome does so at the start codon (labelled codon 1 in Figure 6.4a) which has the sequence AUG (not all AUG codons act as start codons, however, and you will learn later on how the correct start codon of a gene is identified). Once the first tRNA has bound to the mRNA, a second one binds (Figure 6.4a) and the ribosome links the two amino acids, forming what is known as a peptide bond (Figure 6.4b). The first tRNA molecule is then released (Figure 6.4c) and the mRNA moves through the ribosome to present a new codon. tRNAs enter the ribosome one at a time, delivering amino acids which are linked to the growing polypeptide chain (Figure 6.4d). This is called the *elongation* phase of translation. Notice that the first amino acid in the polypeptide has a free amine group ($-NH_2$) so it is said to form the amino terminus or N-terminus of the polypeptide and the last amino acid added to the polypeptide chain has a free carboxyl group ($-COOH$) and forms the carboxy terminus or C-terminus. Polypeptides are always synthesised in the same direction: from the N-terminus towards the C-terminus.

The final event in polypeptide synthesis is *termination* of translation, which is brought about by one of three specific stop codons in the mRNA sequence (UAA, UAG or UGA). When the stop codon is reached, synthesis stops and the completed polypeptide dissociates from the mRNA.

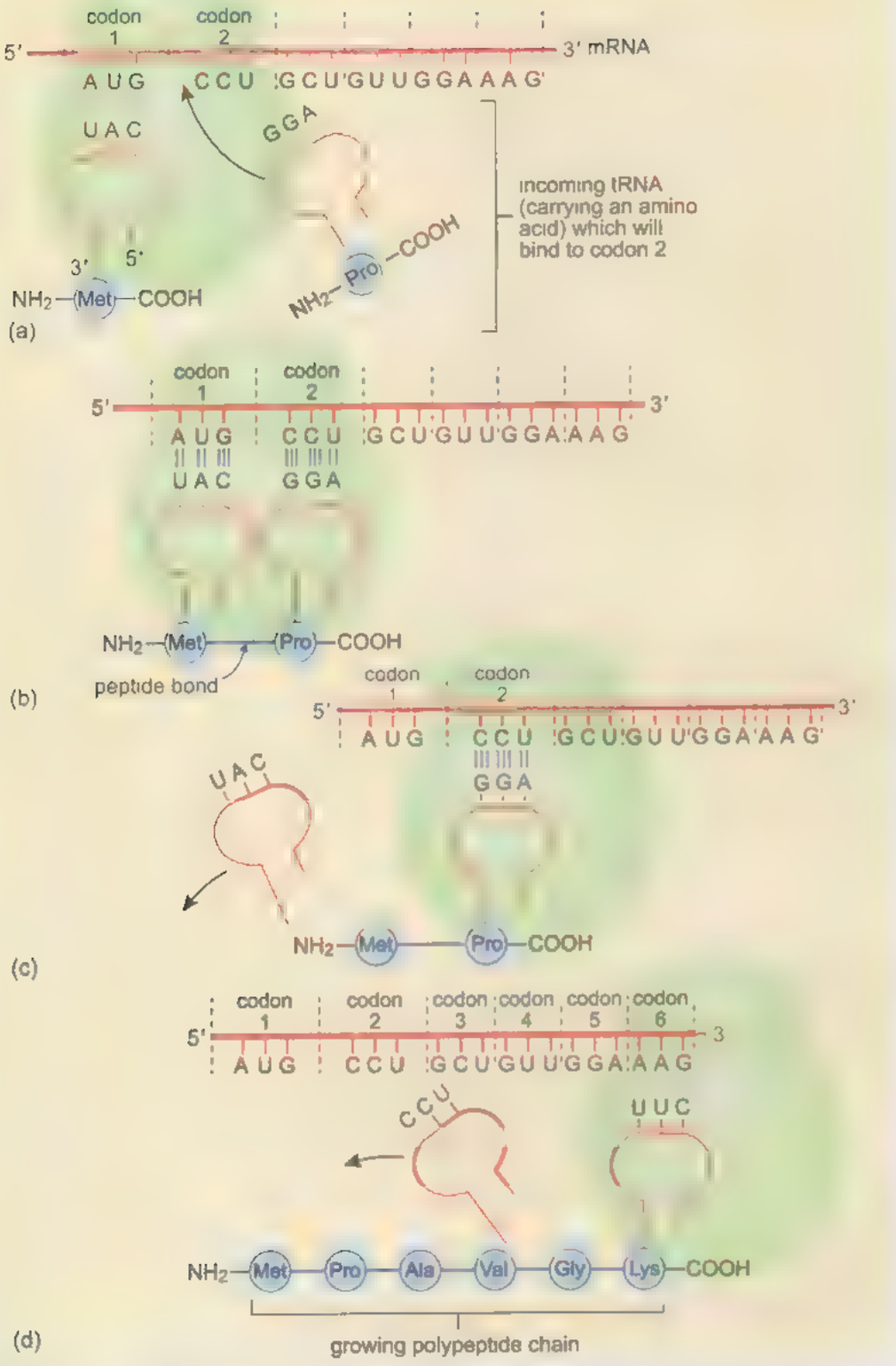



**Figure 6.4** Simplified scheme for translation (a) An mRNA with a tRNA molecule already bound and a second about to bind. The ribosome is shown in green. (b) Two tRNA molecules are now bound to the mRNA, and a peptide bond has formed between the first two amino acids of the polypeptide chain (Met and Pro) (c) The first tRNA molecule is released. (d) A few steps further on in the process the growing polypeptide chain now consists of six amino acids; the sixth amino acid (Lys) corresponds to codon 6 Note that the different components shown are not drawn to scale

## eukaryotes

You are now familiar with the steps by which the linear DNA code of a gene is used to synthesise a polypeptide sequence, but there are important differences in how these processes occur in prokaryotic and eukaryotic cells.

Prokaryotes do not enclose their DNA in a nucleus, so transcription and translation both occur in the cytoplasm, and ribosomes can attach to a new bacterial mRNA and start translating it even while transcription of the mRNA is still being completed. Prokaryotic mRNA is thus immediately ready for translation with no further processing.

In contrast, in eukaryotic cells, the DNA is separated from the cytoplasm by a double membrane, the nuclear envelope. Transcription occurs in the nucleus and the transcribed RNA must then be transported out of the nucleus and into the cytoplasm where the ribosomes are located (either 'free' or attached to the rough ER, Section 3.4.4). Unlike prokaryotic RNA, the **primary RNA transcript** produced in a eukaryotic cell requires extensive **post-transcriptional modification** in the nucleus before it is ready for export to the cytoplasm and translation. One of these modifications is *RNA splicing* (Figure 6.5) to remove some sections of the primary transcript.

- From your understanding of the organisation of the eukaryotic genome (Chapter 5), why is post-transcriptional RNA splicing necessary for eukaryotic but not prokaryotic mRNA?

  Unlike prokaryotic genes, the protein-coding sequences of eukaryotic genes (exons) are interrupted by introns (non-coding sequences). Introns must therefore be removed from the primary RNA transcript to produce a mature mRNA that can be accurately translated into the polypeptide

Splicing removes the introns and joins together the exons to form the mature mRNA (Figure 6.5). The 5′ and 3′ ends of the mature mRNA also undergo modifications, known as RNA *capping* and *polyadenylation* (pronounced 'polly-add-en-elation'), respectively, which will be described in more detail later in this chapter. These terminal modifications prepare the mRNA for export out of the nucleus through the nuclear pores (Section 3.4.3), and into the cytoplasm, where it is finally translated at ribosomes to produce a polypeptide.

Eukaryotes also carry out a wide range of **post-translational modifications** of the translated polypeptide before it becomes a mature functional protein (whereas prokaryotes carry out very few post-translational modifications). Eukaryotic polypeptides may be modified in many different ways by different cellular enzymes, for example, cleavage of the polypeptide into smaller sections, or attachment of additional chemical groups (such as phosphate, carbohydrate or lipid groups) to particular amino acids. The 'signals' for many of these modifications are encoded in the amino acid sequence of the polypeptide itself.

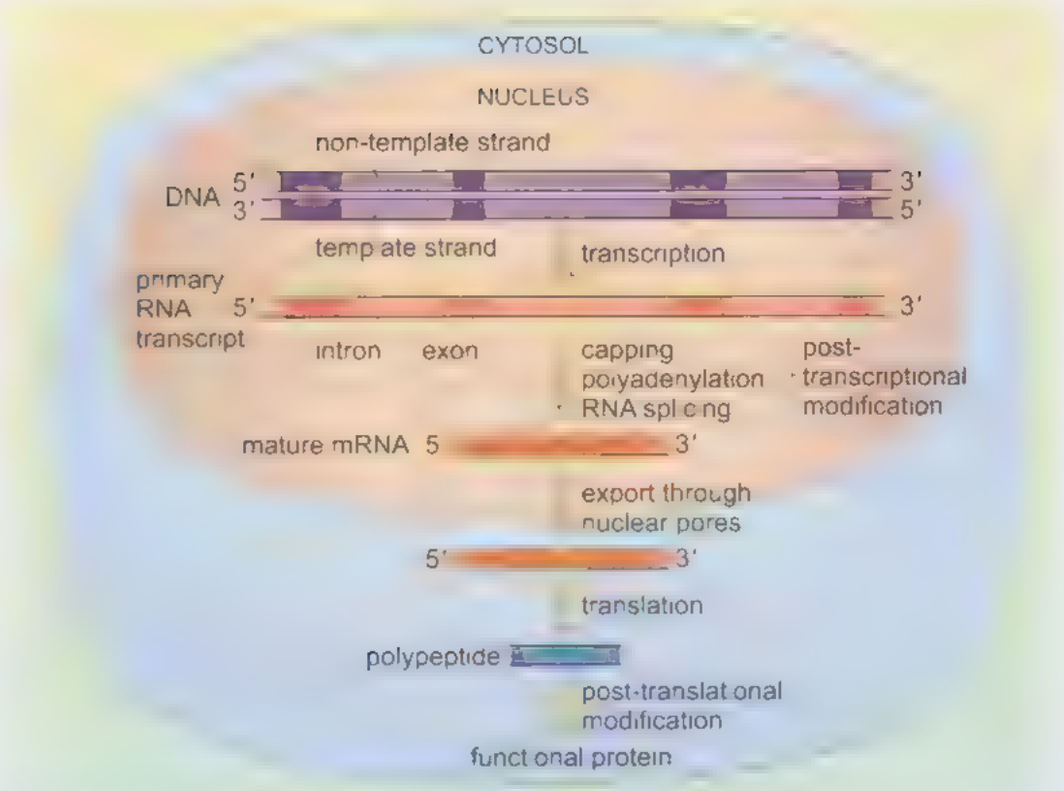



**Figure 6.5** Stages in gene expression in eukaryotes: a summary of events showing the flow of information in a eukaryotic cell from nuclear DNA to a functional protein via primary RNA transcript, mature mRNA and newly synthesised polypeptide. Note the introns – the lighter bands in the DNA and the primary RNA transcript. Introns are removed by splicing of the primary RNA transcript to form a mature mRNA.

- What other types of 'signal' sequences may be present in a polypeptide?

  Some polypeptides contain specific signal sequences that 'target' them to specific parts of the cell, such as the cell membrane or an organelle, or target them for secretion from the cell (Section 3.4.4).

Ultimately, transcription and translation (accompanied in eukaryotes by modification of the mRNA and polypeptide gene products) result in the production of a functional protein, thereby completing the information flow (Figure 6.5).

The control of the processes summarised in this section are examined in more detail in the rest of this chapter. The next two sections look at the control of transcription in prokaryotes and eukaryotes, respectively. In both cases, it is control at the level of transcription that exerts the major influence on the amount and type of gene products present in cells. Transcription can be divided into three main stages. RNA polymerase must first bind to the DNA template at the start of a gene: a process called **initiation**. In the second stage, **elongation**, the RNA polymerase moves along 'reading' the DNA sequence and synthesising a complementary mRNA. In the final, **termination** stage, elongation ceases and the transcript dissociates from the template.

## Summary of Sections 6.1 and 6.2

- A typical cell expresses only a fraction of its gene repertoire at any one time. A multicellular organism with multiple cell types arises because different sets of genes are expressed in different types of cells.
- The sequence of bases in the DNA molecule determines the amino acid sequence of the polypeptide encoded by a protein-coding gene. Each type of polypeptide in a cell is encoded by a specific gene.
- The stages of gene expression are: transcription of DNA into RNA (followed by post-transcriptional modification in eukaryotes) and translation of RNA into polypeptide (followed by post-translational modification in eukaryotes).
- Transcription is the process of RNA synthesis, in which information encoded in one of the strands of DNA is used as a template to produce a single-stranded RNA copy. Accuracy is assured by complementary base pairing in a similar way to DNA replication, except that in RNA, U (not T) pairs with A. RNA polymerase synthesises RNA by linking together ribonucleotides.
- During transcription, RNA polymerase 'reads' the template DNA in a 3′ to 5′ direction, and synthesises the RNA transcript in a 5′ to 3′ direction, such that the transcript is complementary to the DNA template strand.
- Translation is the process of polypeptide synthesis that occurs at the ribosome. A triplet codon of mRNA binds to the corresponding triplet anticodon of a tRNA molecule and the amino acid attached to the tRNA is attached to the previous amino acid in the polypeptide chain via a peptide bond.
- Polypeptides are synthesised from the amino ($-NH_2$) or N-terminus towards the carboxy (–COOH) or C-terminus.
- The genetic code is degenerate; some amino acids are specified by more than one codon.
- In bacteria, transcription and translation occur in the cytoplasm, close to the DNA; while in eukaryotes, transcription occurs in the nucleus and the mature mRNA is exported to the cytoplasm for translation.

## 6.3 Control of prokaryotic gene transcription

Transcription is rather simpler in prokaryotes, so first you will look at an example of the control of bacterial transcription, before moving on to the more complex process of transcription in eukaryotes.

### 6.3.1 Initiation of bacterial transcription

Bacteria have a single type of RNA polymerase that catalyses synthesis of all three main types of RNA (mRNA, rRNA and tRNA). The *E. coli* RNA polymerase is one of the largest proteins in the cell and has five subunits (Figure 6.6). There are four catalytic subunits (two large subunits, β and β′, and two copies of the α subunit, which all together form the *core polymerase* capable of synthesising an RNA chain) and a detachable regulatory subunit

called the *sigma (σ) factor*. The complete, active enzyme assembly can be represented as ααββ′σ.

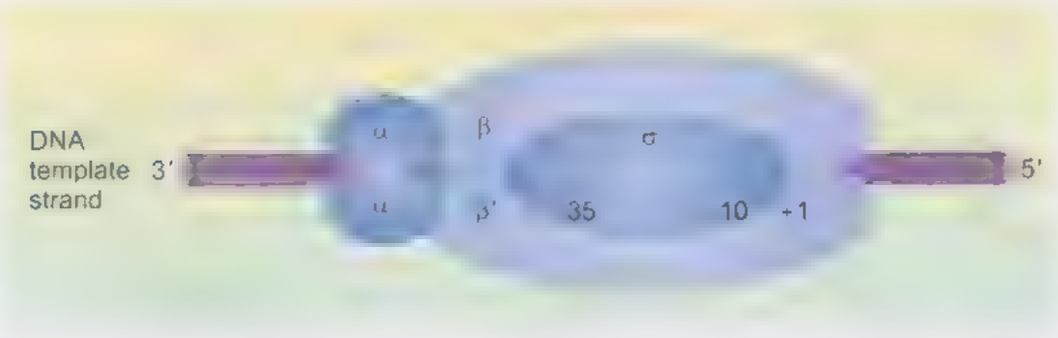



**Figure 6.6** The structure of *E. coli* RNA polymerase attached to DNA (only the DNA template strand is shown). The active form of the enzyme has five subunits, and the complex is designated as ααββ′σ.

Transcriptional initiation begins when RNA polymerase binds to the double-stranded DNA upstream of (before the start of) a gene at a specialised region called the **promoter** (Figure 6.7a). Although the nucleotide sequences of gene promoters differ, some regions of the promoter are very highly conserved; that is, they are always very similar (but not necessarily identical), indicating that this part of the sequence is essential to promoter function. These conserved sequences are known as **consensus sequences**. Prokaryotic gene promoters have a consensus sequence (3′ TATAAT 5′) referred to as the −10 box (or the Pribnow box, after its discoverer) because it is about ten base pairs upstream of the site where transcription starts (Figure 6.7a). Most prokaryotic promoters also have a second consensus sequence (3′ TTGACA 5′) further upstream from the transcription start site, which is known as the −35 box. The RNA polymerase binds to gene promoters in DNA, and not elsewhere, because the σ factor subunit of RNA polymerase specifically recognises the −10 and −35 consensus sequences.

Although all bacterial promoters share these consensus sequences, the rest of the promoter region differs between promoters. Bacterial cells in fact have several different types of σ factors that recognise the consensus sequences in the context of slightly different promoter sequences, so different σ factors have a role in specifying which genes are actively transcribed by the RNA polymerase. The overall promoter sequence determines the 'strength' of the promoter, that is, how often RNA polymerase successfully recognises and binds to the promoter and initiates transcription. In theory, RNA polymerase can bind to all promoters in the genome, so promoters must compete for a restricted pool of RNA polymerase molecules. A 'strong' promoter binds RNA polymerase molecules more often and hence will give a higher rate of transcription, producing more RNA transcript over a period of time, than a 'weak' promoter. The frequency of initiation of transcription at a promoter can also be increased by other DNA binding proteins which associate with specific regulatory DNA sequences in the promoter region. You will encounter this type of regulation in the next section and later on in the chapter.

The promoter sequence itself has polarity; that is, it has to be 'read' in a particular direction. As a result, when the RNA polymerase associates with the promoter, it is appropriately orientatated to transcribe the correct DNA template strand. The **core promoter** (sometimes called the basal promoter), which is present in all prokaryotic genes, includes the binding site for RNA polymerase and the transcription start site (Figure 6.7a).

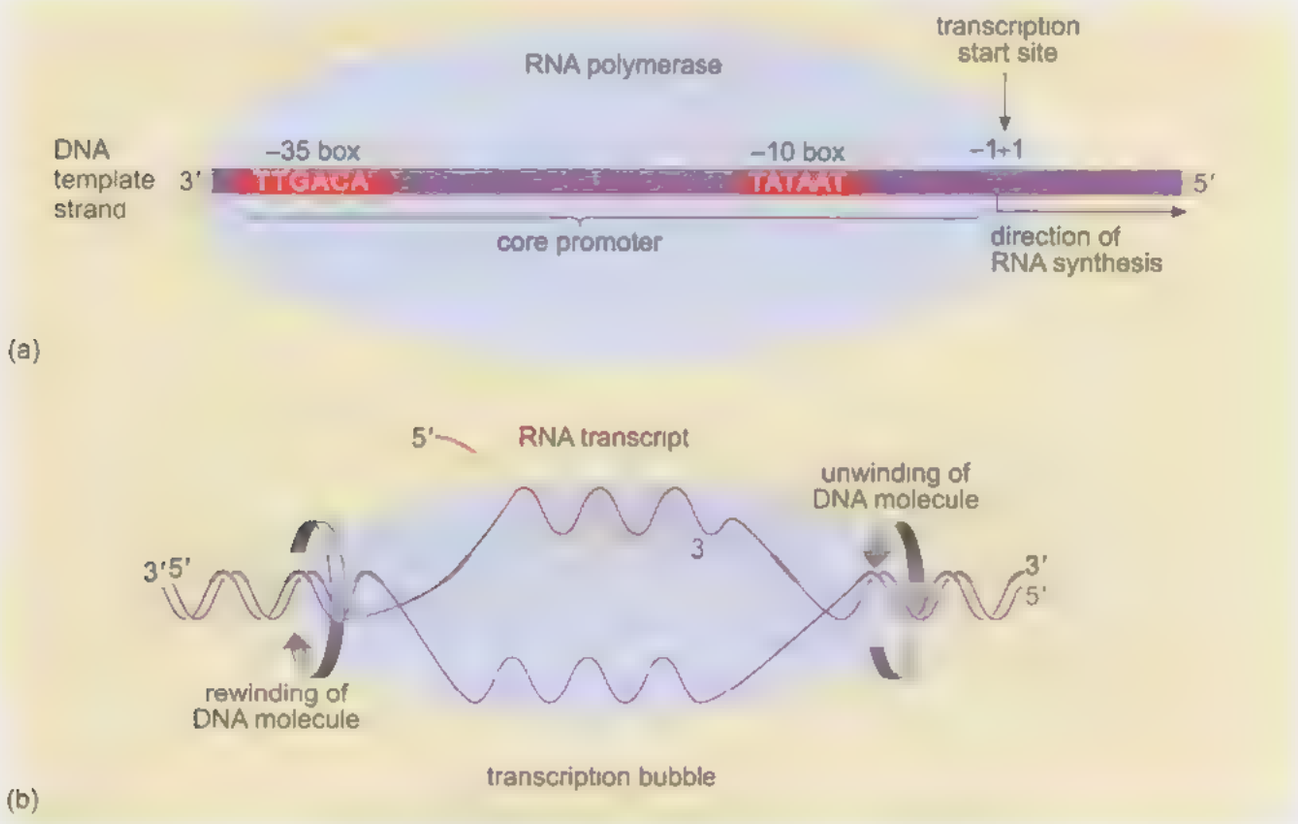



**Figure 6.7** (a) A prokaryote core promoter including the transcription start site (only the template DNA strand is shown). Two specific DNA sequences are recognised by the enzyme RNA polymerase, namely the –35 box and the –10 box. The base at which transcription starts is denoted by +1. The RNA polymerase (blue oval shape) binds to the promoter in double-stranded DNA to form a transcription initiation complex. (b) A transcription bubble. The RNA polymerase unwinds the DNA double helix and transcribes the DNA template strand to synthesise the RNA transcript.

The complex formed once RNA polymerase is tightly bound to the core promoter is called a **transcription initiation complex**. Once bound, the DNA–protein complex converts from a *closed* complex (Figure 6.7a) into an *open* complex (Figure 6.7b) by the unwinding of a short stretch of the DNA double helix and separation of the two strands to expose 15–20 unpaired nucleotides on each strand. This is sometimes referred to as a *transcription bubble*.

At this point the σ factor is released from the RNA polymerase and RNA chain elongation begins. As the polymerase moves off, the promoter region is exposed again. This 'promoter clearance' signifies the end of initiation. The exposed promoter can immediately be bound by another molecule of RNA polymerase which initiates a new transcript. Transcription of the RNA actually starts at a specific base pair in the DNA template strand. This is called the

**transcription start site** and is defined as position +1 (Figure 6.7a). The nucleotide immediately preceding it is defined as position −1; there is no position 0. The RNA polymerase advances along the double helix, unwinding the DNA and rewinding it behind, such that the transcription bubble moves along the DNA, enabling one strand to act as the template for the synthesis of the RNA transcript. RNA synthesis continues, at about 50 nucleotides per second for bacterial RNA polymerase, until the enzyme encounters a *terminator* sequence in the DNA (described in Section 6.3.4), at which point the RNA transcript is released, and the RNA polymerase detaches from the DNA template. The freed polymerase reassociates with another σ factor and can bind to another gene promoter to start initiation of a new transcript.

## 6.3.2 Transcription is controlled by the binding of

The initiation step of transcription is the main point at which the level of gene expression is controlled in both prokaryotes and eukaryotes. As well as the core promoter described above (which includes the site where RNA polymerase binds and the transcription start site) almost all prokaryotic and eukaryotic genes also have several upstream (and occasionally downstream) **regulatory DNA elements**. These are short consensus sequences in DNA that are specifically recognised and bound by DNA binding proteins called **transcription factors**. These factors either promote the recruitment of RNA polymerase to the promoter, thereby increasing transcription of the gene (in which case they are called transcriptional *activators*), or they repress recruitment of the RNA polymerase and decrease transcription of the gene (transcriptional *repressors*).

- What other kinds of protein directly bind to DNA?

  Some examples of DNA binding proteins are RNA polymerases, histones and DNA replication and repair enzymes (Chapter 5).

Transcription factor proteins have two characteristic domains. The *DNA binding domain* interacts with the sugar–phosphate backbone and the bases of the DNA strands in such a way that the protein fits tightly into grooves of the DNA double helix. The second domain, the *transactivation domain*, interacts with other transcription factors, or with proteins bound at the core promoter, to regulate formation of the transcription initiation complex. There are many types of transcription factors, but because they all need to bind to DNA they tend to share a few common structural features called **DNA binding motifs**. The precise amino acid sequence of the DNA binding motif forms a binding site of a specific shape that determines which particular DNA sequence is recognised by the protein factor. Transcription factors can be grouped into a few families on the basis of their common DNA binding motifs, which have been highly conserved through evolution, and are present in a range of proteins from very different organisms. Three typical DNA binding motifs, namely the *helix–turn–helix motif*, the *zinc finger*, and the *leucine zipper*, are illustrated in Figure 6.8.

- What is a common feature of the interaction between the three DNA binding motifs and the DNA helix shown Figure 6.8?

  All three of these DNA binding motifs interact with the major groove of the DNA helix (Section 5.2.1)

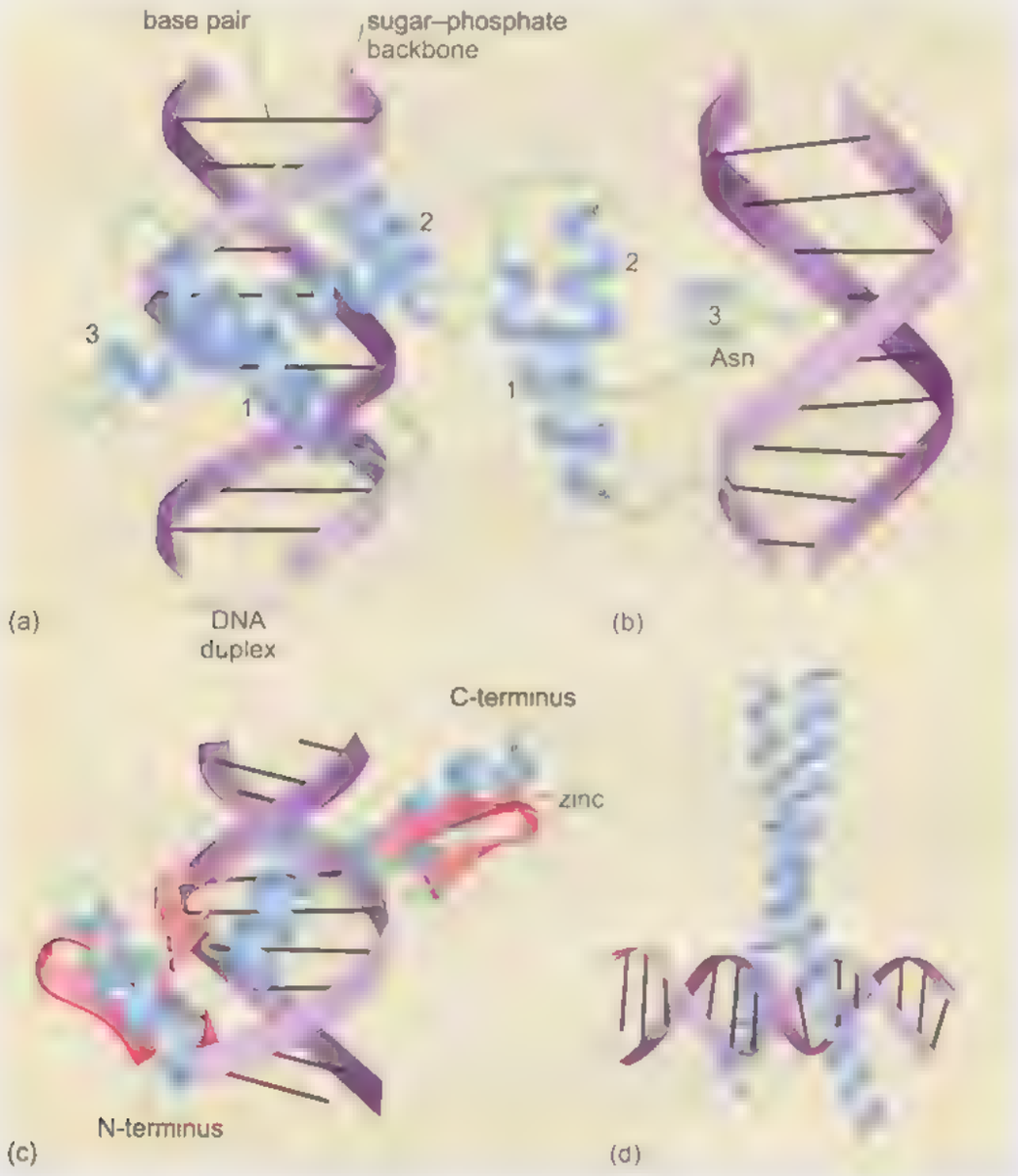



**Figure 6.8** The structural motifs of three major types of DNA binding proteins (a) and (b) Frontal and lateral (side) views, respectively, of the helix turn helix protein motif, which consists of three α-helices (in blue) that fit into the major groove of the DNA helix. The α-helix and other forms of protein secondary structure are described in more detail in Book 2, Chapter 1 (c) A cluster of three zinc finger motifs, so called because they contain small finger-like projections each consisting of an α-helix (blue) and a β-sheet (red) held together by a zinc atom (orange). Zinc fingers are rare in prokaryotes, but are very common in eukaryotes (d) A leucine zipper motif formed from two α-helices, belonging to separate protein molecules that together 'grip' the DNA molecule rather like a clothes peg.

The next section describes an example of gene control in *E. coli* that involves two transcription factors from the helix–turn–helix structural family. One is a transcriptional repressor called the *lac* repressor and the other is a transcriptional activator called the catabolite activator protein (CAP).

### 6.3.3 The bacterial operon

Bacteria live in a dynamic environment in which the levels of nutrients, metabolites and other molecules are constantly changing, so the ability to rapidly alter the level of proteins required to process these molecules is essential for survival. The bacterial genome is organised in such a way that genes encoding proteins involved in the same biochemical pathway, for example the enzymes and transporter proteins responsible for the import and catabolism of nutrients, are often grouped together in units called operons. An operon is expressed as a single polycistronic mRNA transcript, controlled by a single promoter (Section 5.7.1). Altogether, there are about 600 operons in the *E. coli* genome.

The ***lac* operon** of *E. coli* is perhaps the best-studied bacterial operon. It was first described by François Jacob and Jacques Monod, who together won the Nobel Prize for Physiology in 1965 for their contribution to the understanding of gene control. The *lac* operon coordinates the expression of three genes required for the utilisation of lactose (the disaccharide that is the major sugar found in milk) as a source of energy. These three genes are known as: *lacZ* (which encodes β-galactosidase, an enzyme that breaks down lactose into glucose and galactose), *lacY* (which encodes β-galactoside permease, a membrane protein that transports lactose into the cell); and *lacA* (which encodes β-galactoside transacetylase, an enzyme whose function is currently unclear).

*E. coli* cells growing in a laboratory culture medium that contains glucose as the only source of carbon and energy have very low levels of β-galactosidase and β-galactoside permease activity. If the cells are switched to a medium containing lactose instead of glucose, however, the level of expression of the β-galactosidase and β-galactoside permease proteins increases within about 20 minutes. If the bacteria are then switched back from the lactose-containing medium to a glucose medium, β-galactosidase and β-galactoside permease expression is shut off, or *repressed*, within minutes of the changeover.

How are these rapid changes in gene expression triggered by the presence or absence of lactose? The answer lies in the binding of specific transcription factors to regulatory DNA elements in the *lac* operon. A region called the *lac* control region lies immediately upstream of the three *lac* genes (*lacZ*, *lacY* and *lacA*). The *lac* control region consists of a promoter region *P*, and two regulatory DNA elements called the activator site (*AS*) and the operator (*O*) (Figure 6.9a).

RNA polymerase is only able to bind to the promoter (*P*) and transcribe the three *lac* genes into a single polycistronic mRNA when lactose is the main energy source available to the bacterium. The bacterium has two mechanisms that impose this regulatory control. Each of them employs a transcription

factor protein – in one case a *repressor* and in the other case an *activator* – that binds to one of the regulatory DNA elements in the *lac* control region.

### The *lac* transcriptional repressor

The first mechanism of regulating the *lac* operon employs a transcriptional repressor called the *lac* repressor. The *lac* repressor is a protein encoded by the *lacI* gene which is situated nearby the *lac* operon. The *lac* repressor protein is expressed *constitutively*, meaning that it is always 'on', and its expression remains at a constant level in the cell. When the cell is growing in a medium containing only glucose (no lactose), the *lac* repressor binds tightly to the *lac* operator DNA sequence (Figure 6.9b) and blocks the binding of RNA polymerase to the promoter. Consequently, there is no transcription from the *lac* operon and the cell contains very little β-galactosidase and β-galactoside permease.

When cells are switched to a medium containing lactose, the lactose molecules enter the cell and bind to the *lac* repressor protein, causing it to undergo a conformational change in the part of the repressor protein molecule that is required for binding to DNA. In biochemistry this is known as *allosteric regulation*, the regulation of a protein's activity by the binding of an effector molecule (in this case lactose) at a site other than the protein's active site (in this case the active site is the repressor's DNA binding site). You will learn more about this type of regulation in Book 2. The conformationally altered *lac* repressor is no longer able to bind to the operator (Figure 6.9c), so the promoter becomes available for RNA polymerase binding.

The control imposed by the *lac* repressor is known as **negative inducible regulation**, because gene expression is induced only when the *lac* repressor protein is *inactivated* by binding to lactose.

- What would be the consequences for *lacZ* gene expression in a cell with a *lacI* gene mutation that resulted in *lac* repressor protein unable to bind to lactose?

  The cell would be unable to express *lacZ*, even in the presence of lactose. If the *lac* repressor could not bind to lactose, the repressor would always remain bound to the operator, and RNA polymerase would be unable to transcribe any of the *lac* operon genes.

However, transcription of the *lac* genes also depends on a second mechanism of regulation that employs a transcriptional *activator* protein.

### The CAP transcriptional activator

*E. coli* 'prefers' to metabolise glucose, even if other sugars are present in the medium, so the expression of the enzymes necessary to metabolise other types of sugars are not required until all the available glucose is used up.

The signal that detects the presence of glucose is cyclic adenosine monophosphate (cAMP), a small signalling molecule whose level in the cell is inversely proportional to that of glucose, in other words, when glucose is high in the cells, cAMP levels are low, and vice versa. In low glucose conditions cAMP molecules are available to bind to an *E. coli* transcriptional activator

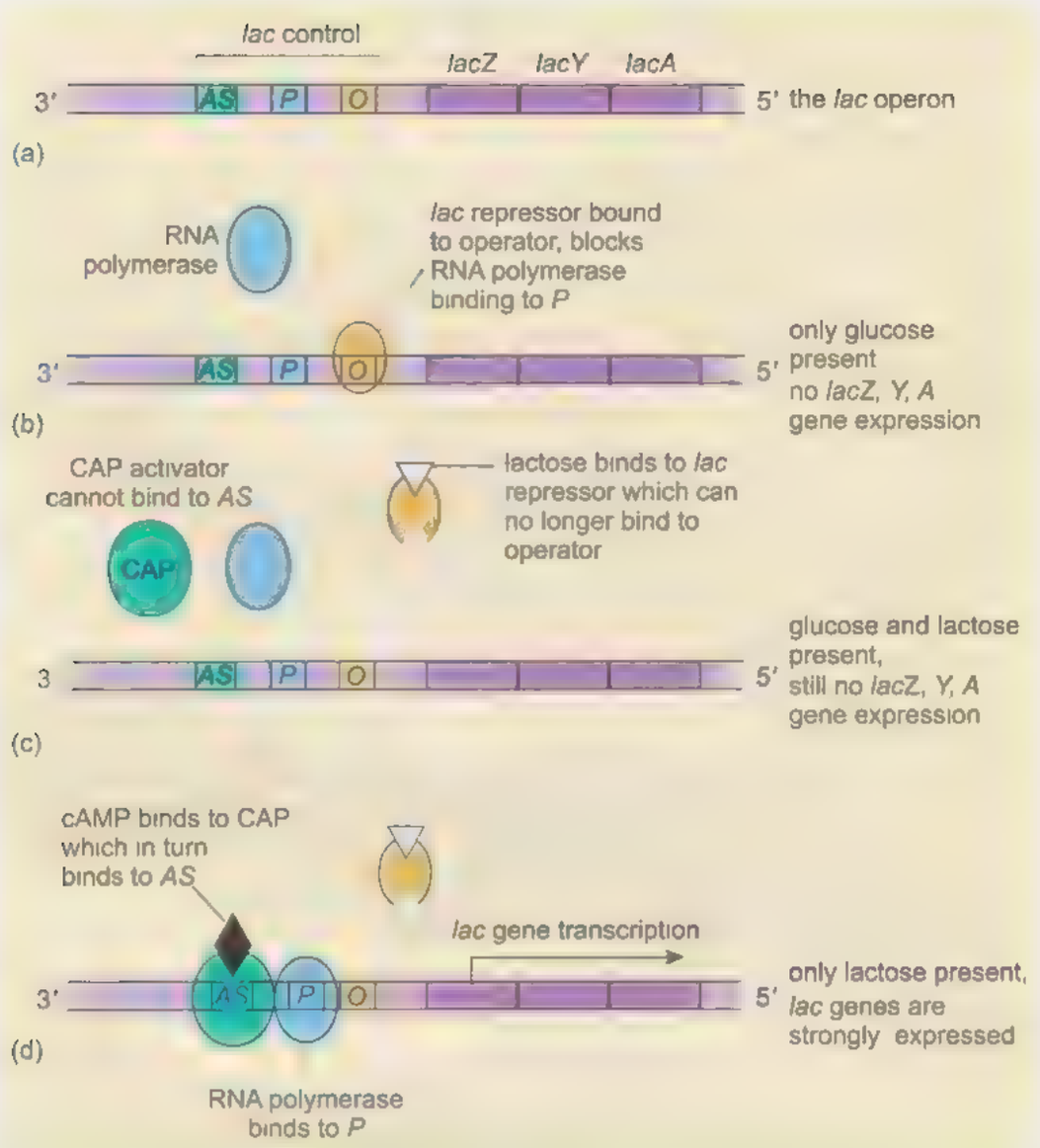



**Figure 6.9** The control of transcription at the *lac* operon in the presence of lactose and glucose. (a) The *lac* operon in the bacterium *E. coli* (see description in the text). Only the template DNA strand is shown and the −10 and −35 boxes are not shown. (b) In the absence of lactose (only glucose is present), the *lac* repressor protein (yellow oval) binds to the operator sequence (*O*) of the *lac* operon, and blocks the interaction of RNA polymerase (blue oval) with the promoter (*P*). (c) When some lactose is present in the growth medium, it binds to the *lac* repressor protein, inducing a conformational change such that the *lac* repressor protein is no longer able to bind to the operator sequence. However, if glucose is also present, cAMP level remains low so catabolite activator protein (CAP, green circle) cannot bind to the DNA. (d) When only lactose is present (no glucose), cAMP levels rise allowing CAP to bind to the activator sequence (*AS*), which facilitates RNA polymerase binding to the promoter (*P*) and transcription of the *lac* genes is initiated.

protein called **catabolite activator protein (CAP)** which thereby undergoes an allosteric conformational change allowing it to bind to the *lac* operon at the activator site *AS* (Figure 6.9a). The binding of CAP to the *AS* is essential for efficient transcription of the *lac* operon because CAP facilitates the binding of

RNA polymerase to the promoter. Without CAP binding to *AS* there is very little transcription, even if the *lac* repressor is not bound to the operator (Figure 6.9c). As long as some glucose is available to the cell, cAMP levels remain low and the CAP activator cannot bind to the activator site (*AS*). Thus RNA polymerase does not bind to the promoter and the *lac* operon is not transcribed, even though there is also some lactose present in the growth medium.

Only when the glucose is depleted does the cAMP level start to rise and alter the CAP activator so that it binds the activator site (*AS*), assists the RNA polymerase to bind to the promoter and initiates transcription. The *lac* genes are expressed (Figure 6.9d) and the β-galactosidase and β-galactoside permease proteins very rapidly reach a high level in the cell. Lactose can then be imported into the cell (facilitated by β-galactoside permease) and hydrolysed by β-galactosidase to release galactose and glucose which are further metabolised to derive energy (Book 2, Chapter 3).

The type of control imposed by the CAP activator is called **positive inducible regulation**, because *lac* gene expression is induced only when the CAP activator binds to an inducer, in this case cAMP.

### 6.3.4 Termination of transcription of bacterial genes

Once initiated, RNA transcript elongation progresses until the RNA polymerase encounters a **transcription terminator** sequence in the RNA. The terminator will be located after the protein coding region in an mRNA, and in *E. coli*, transcription terminates at the end of the last protein-coding region of an operon.

Bacterial RNAs have two types of transcriptional terminator. One type is a binding site for an RNA binding protein called Rho factor which is a helicase (Section 5.3), that unwinds the RNA from the DNA template and disrupts the complementary base pairing between them. This terminates transcription and releases the RNA transcript. The other type of terminator is an *inverted repeat* sequence which, as the RNA is transcribed, folds back on itself to form a hairpin-shaped loop by complementary base pairing *within* the RNA strand (Figure 6.10). The hairpin structure destabilises the interaction between the RNA polymerase and the DNA–RNA hybrid, thus terminating transcription.

Premature (early) termination of transcription is also commonly used by bacteria as a mechanism of downregulating gene expression, this is known as **attenuation**. Attenuators are terminator sequences that form hairpin-shaped structures close to the 5′ end of the RNA, often in response to an environmental signal. The hairpin prevents transcription of the full length RNA. For example, in the presence of high levels of the amino acid tryptophan, an attenuator hairpin forms close to the 5′ end of the polycistronic mRNA transcribed from the *trp* operon (which encodes the components for production of tryptophan). This prevents transcription of the full length *trp* mRNA, so no protein translation can occur, and no tryptophan is synthesised. The *trp* mRNA is only fully transcribed and translated when tryptophan levels are low so the attenuator hairpin is not present and the bacteria can synthesise the tryptophan it requires.

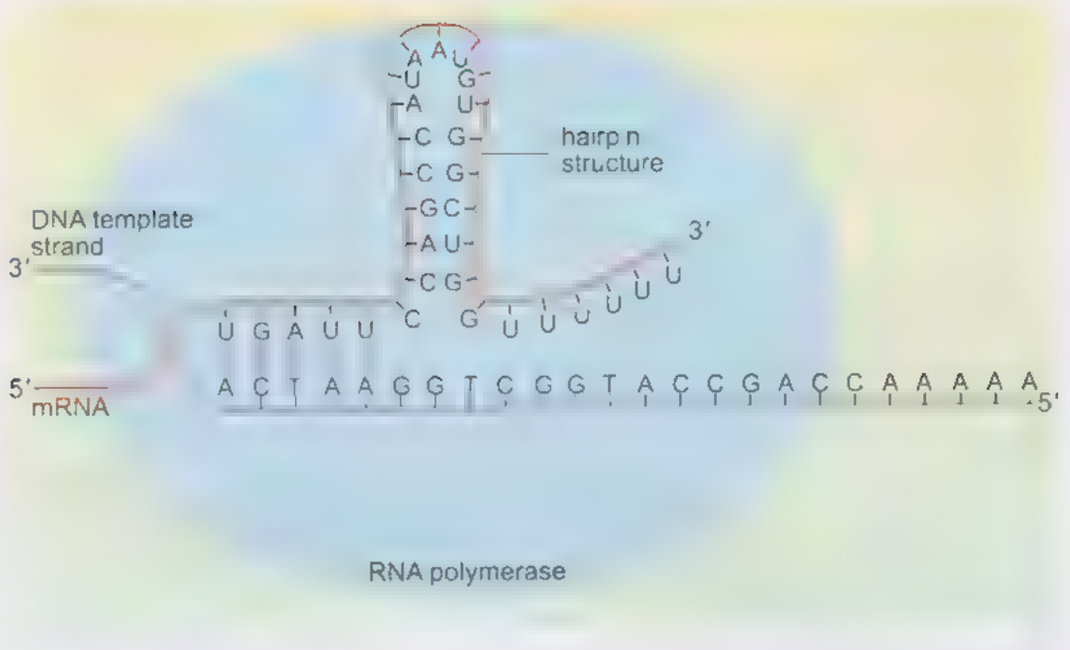



**Figure 6.10** Termination of mRNA synthesis by the formation of a hairpin loop structure within a transcription terminator sequence in the transcribed RNA strand

## Summary of Section 6.3

- Gene transcription starts at a precise place in the gene called the transcription start site.
- Bacterial RNA polymerase has five subunits: two large subunits (β and β′), two identical α subunits, and the regulatory σ factor.
- RNA polymerase binds to a special sequence of bases in the core promoter, slightly upstream of the transcription start site. Bacterial promoters have two consensus sequences where the polymerase binds, these are the −10 and the −35 boxes.
- The σ factor subunit is responsible for recognising the −10 and −35 box consensus sequences.
- In bacteria, genes for related functions are often grouped together in units called operons. Genes in an operon are regulated together, and are under the control of the same promoter.
- Transcription factors (activators and repressors) regulate bacterial operons by binding to specific sites in the promoter (e.g. the operator and activator sites of the *lac* operon).
- In the *lac* operon, the *lac* repressor and the CAP–cAMP activator complex regulate transcription by binding to the operator site and the activator site, respectively. In the absence of lactose, the *lac* repressor binds to the operator and inhibits transcription (negative inducible regulation). In the presence of lactose the *lac* repressor can no longer bind to the DNA. Binding of CAP–cAMP to the activator site activates transcription (positive inducible regulation) once no glucose is present.
- In bacteria, termination of transcription occurs at terminator sequences where a terminator protein, Rho helicase, separates the RNA transcript

from the DNA, or where disruptive secondary structures (hairpin loops) form in the transcribed mRNA. Attenuation, premature termination resulting in truncated mRNA molecules that are not translated, is a mechanism by which transcription is downregulated in bacteria.

## 6.4 Control of eukaryotic gene transcription

Gene regulation in eukaryotes is generally far more complex than in bacteria. In a multicellular organism, as well as controlling the continual changes that each cell must make in response to its environment, gene regulation is also responsible for the programme of cell differentiation that generates multiple cell types. You will learn more about differentiation in Book 3.

Gene regulation in an individual cell is also modulated by many different types of signals that originate from surrounding cells or from elsewhere in the organism. For example, all multicellular organisms produce **growth factors** and **hormones** (phytohormones in plants) – signalling molecules that bind to cells carrying appropriate receptor proteins. These molecules activate a signal transduction mechanism inside the cell that ultimately alters gene expression and leads to cell type-specific responses, such as a change in metabolism. You will learn more about signal transduction mechanisms in Book 2, Chapter 4.

Eukaryotic gene expression is regulated at multiple stages and the complexities are well beyond the scope of this module, but in this section, the main mechanisms that control eukaryotic gene transcription will be discussed. The regulation of post-transcriptional mRNA modification and protein translation will be briefly described in the remainder of the chapter.

Transcription in eukaryotes differs in three important ways from that in prokaryotes.

(a) Eukaryotic polymerases cannot bind directly to DNA because they do not contain a subunit equivalent to the prokaryotic σ factor which facilitates polymerase binding to the core promoter (Section 6.3.1). In eukaryotes, this function is instead carried out by a number of **general transcription factors** which must assemble with the RNA polymerase at the promoter before transcription can be initiated.

(b) Prokaryotic DNA is essentially 'naked' and freely accessible to the transcription proteins, but in higher eukaryotes the nuclear DNA is packaged with histone proteins to form densely packed chromatin (Section 3.4.3). Before transcription can take place, eukaryotic DNA must therefore be made accessible to the transcription machinery.

(c) The level of transcription of eukaryotic genes is regulated by multiple **specific transcription factors** that bind to regulatory DNA sequences spaced at intervals along the DNA, some of which may be thousands of base pairs away from the gene that they influence. These regulatory transcription factors work by facilitating the processes described in (a) and (b) above.

## 6.4.1 The eukaryotic transcription initiation complex

There are several types of eukaryotic RNA polymerases, which transcribe different types of genes. The transcription products of the three most important eukaryotic polymerases are summarised in Table 6.1. RNA polymerase II is the polymerase that transcribes most of the protein-coding genes, and the mechanisms described in this section relate to this enzyme.

**Table 6.1** RNA polymerases in eukaryotic cells.

| Type of RNA polymerase | Transcription product |
|---|---|
| RNA polymerase I | rRNA |
| RNA polymerase II | mRNA and most small nuclear RNA (snRNA) and microRNA |
| RNA polymerase III | tRNA, small ribosomal RNA, and other small RNAs found in the nucleus and cytosol |

RNA polymerase II typically contains 10–12 subunits. Three of these are core subunits with sequence *homology* (i.e. their protein sequences show similarity) to the *E. coli* core polymerase subunits ($\alpha\alpha\beta\beta'$, Section 6.3.1). This suggests that the structure of the core polymerase subunits is highly conserved across all species. The exact roles of some of the smaller subunits are poorly understood. Some of them are common to all types of eukaryotic polymerases while others are specific to a particular type of RNA polymerase.

The core promoter of a protein-coding eukaryotic gene typically contains a highly conserved sequence called the **TATA box** usually between −25 and −35 bp upstream of the transcription start site (Figure 6.11), so called because the most frequently occurring bases in the TATA box are the six-base sequence TATAAA.

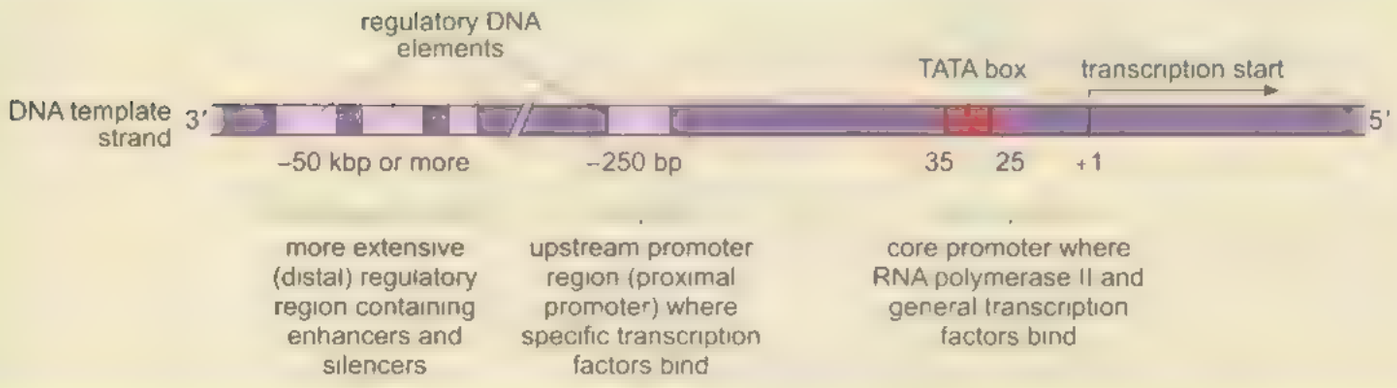



**Figure 6.11** The transcription control elements that control protein-coding gene expression in a typical eukaryotic gene. Only the template DNA strand is shown. RNA polymerase II binds to the core promoter, which includes the TATA box and the transcription start site, while regulatory transcription factors bind to the proximal promoter region about 250 bp upstream of the transcription start site, or to more distant regulatory elements, enhancers and silencers, which may be many kilobase pairs (kbp) distant from the transcription start site.

- How does the TATA box sequence compare with the consensus sequence of the −10 box of the prokaryotic promoter shown in Figure 6.7?

There is homology (sequence similarity) between these two consensus sequences.

This suggests that the sequence has been tightly conserved during evolution as the binding site for the RNA polymerase complex in all organisms. RNA polymerase II and a group of proteins called *general transcription factors* (to distinguish them from the specific transcription factors that regulate the level of transcription and which will be described later on) must bind to the core promoter to form the transcription initiation complex. The first stage of transcriptional initiation is the recognition of the TATA box by the general transcription factor TFIID (transcription factor IID, where II is the Roman numeral two). This is a protein complex consisting of a TATA binding protein (TBP) and a number of TBP-associated factors (TAFs) (Figure 6.12).

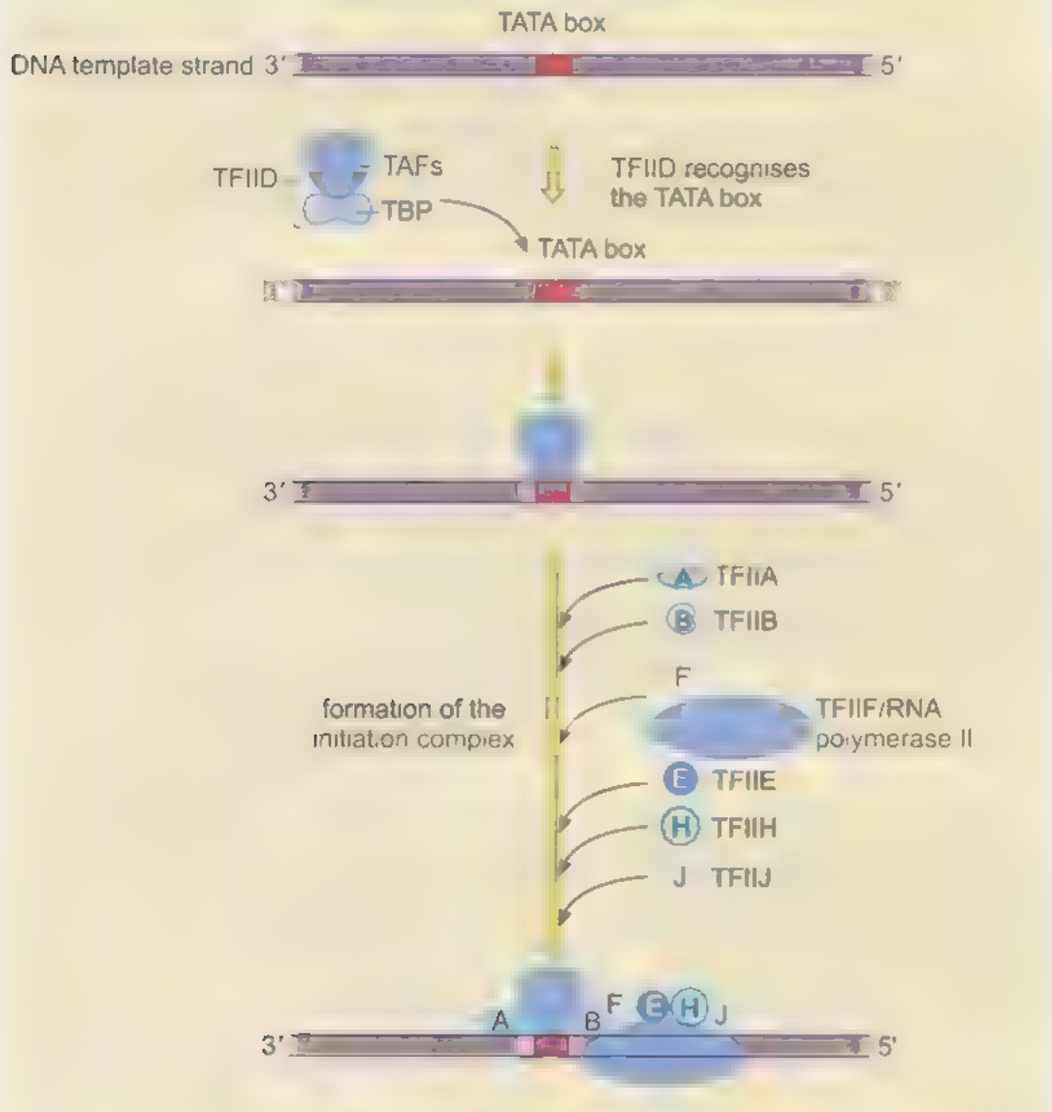



**Figure 6.12** Assembly of the RNA polymerase II transcription initiation complex on a TATA-containing promoter (see description in the text). Only the template DNA strand is shown.

A number of other general transcription factors (you don't need to memorise all these) are then 'recruited' to the promoter and facilitate the binding of RNA polymerase II and separation of the strands of DNA, thereby allowing the transcription bubble to form. The RNA polymerase II bound to the promoter becomes phosphorylated, causing it to undergo a conformational change which releases it from the initiation complex to move along the DNA and start elongating the new mRNA chain. Most of the general transcription factors are released from the DNA at this point so that they are available to initiate transcription at another site.

You may be surprised to learn that the transcription initiation complex, including the polymerase, consists of at least 40 polypeptides at this stage.

To summarise, the three major roles of general transcription factors are

- positioning the RNA polymerase correctly at the promoter
- separation of the two strands of DNA to allow transcription to begin
- release of the RNA polymerase from the initiation complex to start transcription.

## expression

The level of expression of a eukaryotic gene is, as in prokaryotes, mainly determined by transcriptional activator and repressor proteins that regulate transcriptional initiation (the binding of the polymerase to the core promoter). Regulation of initiation in a eukaryotic gene is far more complex than in a bacterial operon.

Some eukaryotic genes, often called **housekeeping genes**, are expressed *constitutively* (at a relatively constant level) in most cell types. These tend to be genes involved in some of the basic functions necessary for the maintenance of the cell – for example, genes that encode ribosome components and many of the metabolic enzymes. Most genes, however, are *inducible*; meaning that they are only expressed in particular cell types at particular times, or in response to signals such as growth factors.

Transcriptional initiation is controlled by regulatory DNA elements that are recognised and bound by specific transcription factors in the appropriate circumstances. Transcription factors can be constitutively active or they may require activation by intracellular or extracellular signals, for example, the 'response factors' shown in Table 6.2 that are activated in response to growth factors circulating in the blood serum or to cellular stresses, such as increased temperature. Some transcription factors are ubiquitous, that is, they are expressed in all types of cell, while others are cell-type specific. Most genes are only expressed when a particular combination of several different transcription factors binds to the appropriate regulatory elements. To illustrate the diversity of these transcription factors and the regulatory DNA elements to which they bind, a few examples are given in Table 6.2.

**Table 6.2** Examples of specific transcription factors and the regulatory DNA sequences they recognise. (N = any nucleotide)

| Factor | Regulatory DNA sequence recognised | Comments |
|---|---|---|
| *Constitutive transcription factors* | | |
| Sp1 | 5′ GGGCGG 3′ | ubiquitous |
| Oct-1 | 5′ ATGCAAAT 3′ | ubiquitous |
| *Response factors* | | |
| serum response factor | 5′ CCATATTAGG 3′ | activated by growth factors in serum |
| heat shock factor | 5′ CNNGAANNTCCNNG 3′ | activated by heat shock |
| *Cell-specific factors* | | |
| GATA-1 | 5′ GATA 3′ | specific to erythroid cells (which give rise to red blood cells) |
| MyoD1 | 5′ CANNG 3′ | specific to myoblast cells (which give rise to skeletal muscle) |

As described in the previous section, eukaryotic protein-coding genes have a core promoter where the general transcription factors and the polymerase assemble. Most also have an upstream promoter region (sometimes called the proximal promoter) about 250 bp upstream of the TATA box (Figure 6.11) which contains binding sites for specific transcription factors (which can be inducers or repressors) that are necessary to control the level of gene transcription.

Many eukaryotic genes also have a much larger and more variable 'distal promoter region' containing regulatory elements. These may be up to several thousand bases upstream of the transcription start site (Figure 6.11). In addition, there are regulatory DNA elements called *enhancers* and *silencers* that may be situated quite close to the promoter but can be much further from the gene, sometimes up to 50–100 kbp distant. Unlike promoter regulatory elements, enhancers and silencers can work in either orientation and may affect more than one gene. **Enhancers** are bound by transcriptional activator proteins which stimulate gene transcription. **Silencers** bind transcriptional repressor proteins and have the opposite effect of reducing gene transcription. Enhancers and silencers are very common in eukaryotes, but rare in bacteria. The first enhancer to be discovered was that encoded by simian virus 40 (SV40), a virus (with a DNA genome) that infects eukaryotic cells. The enhancer was identified by joining up regions of DNA to core promoter sequences to see which combinations stimulate transcription of a 'reporter' gene (Box 6.1). The SV40 enhancer has been found to be capable of activating the transcription of many eukaryotic genes, and is often used by researchers as a component of plasmids employed to express recombinant proteins in cultured eukaryotic cells. Another example of an enhancer relates to the regulation of genes encoding antibodies (immunoglobulins). These genes contain an enhancer that can stimulate transcription, but is active only in B cells (the type of immune system cell that makes antibodies).

## Box 6.1 Reporter genes for analysis of gene expression

In order to study the regulatory potential of DNA sequences in a gene promoter region, sections of the DNA sequence can be isolated, linked to an easily detectable reporter gene and reintroduced into cultured cells, or even animals or plants. The level of expression of the reporter gene in the cell is an indication of whether the DNA sequence contains elements that can act as constitutive or regulatory promoter elements.

Certain genes are routinely used as reporters because the characteristics they confer on the cells expressing them are easily identified and measured. Common reporter genes are the green fluorescent protein (GFP) gene isolated from jellyfish, which causes the cells that express it to glow green when exposed to ultraviolet light (Figure 6.13), and the gene encoding luciferase, an enzyme that oxidises a pigment called luciferin to release light, which can be measured using a luminometer. Another common reporter is β-galactosidase, an enzyme that cleaves the sugar lactose (Section 6.3.3). A colourless modified form of lactose called Xgal is added to the cells, and if β-galactosidase is expressed, it converts the Xgal into a blue product (Box 5.2) that can be detected by absorbance of light using a spectrophotometer.

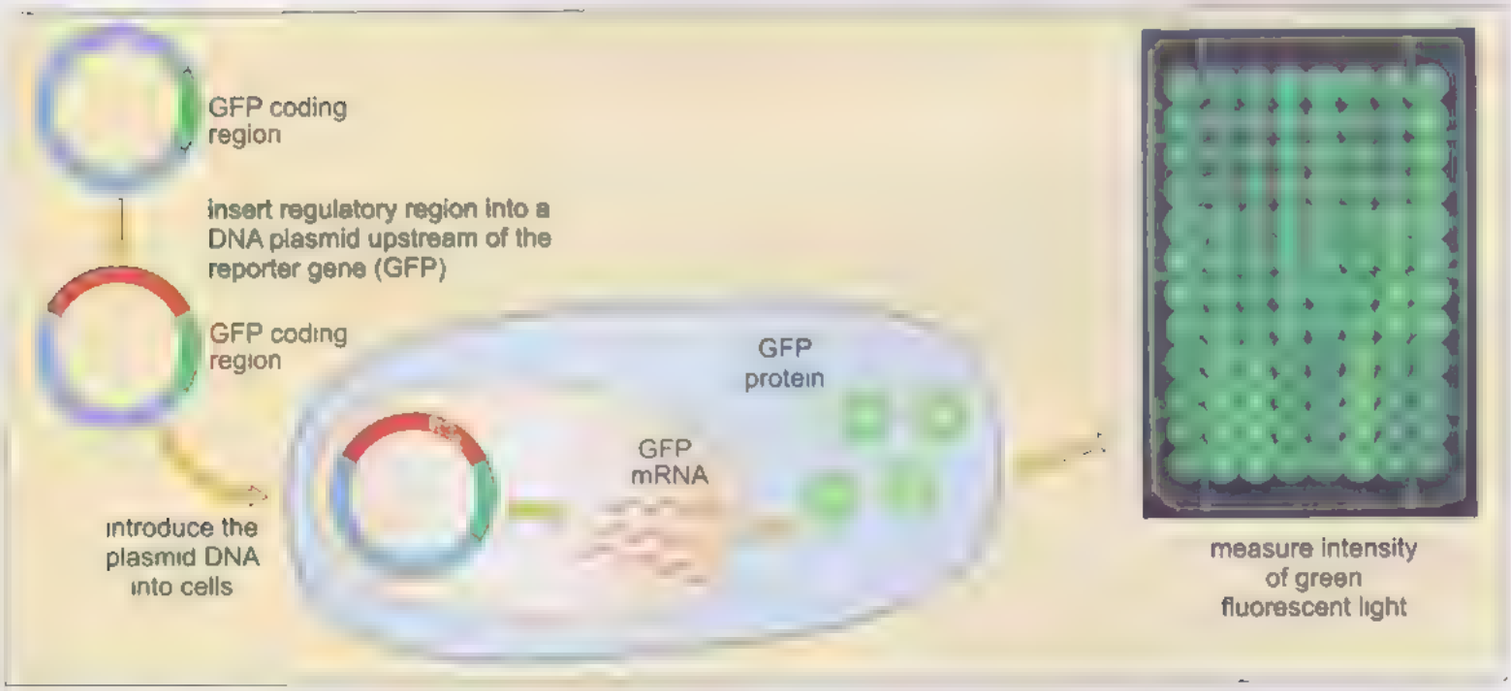



**Figure 6.13** Reporter gene assay for gene expression. A plasmid (Box 5.2) carrying regulatory DNA sequences fused to the coding region of the green fluorescent protein (GFP) gene is introduced into cells which transcribe and translate the GFP gene only if the regulatory DNA sequence is active in the host cells, i.e. if the appropriate regulatory transcription factors are present. The amount of GFP protein produced therefore reflects the activity of the regulatory DNA sequence and can be measured by exposing the cells to UV light and using a fluorescent microscope or a fluorimeter to measure the intensity of the green fluorescent light emitted by the cells. Shown here are multiple cell samples fluorescing in a 96-well plate.

### 6.4.3 Coordinating the initiation of transcription

How do these various regulatory DNA sequences and transcription factors work together to regulate transcription? In many cases, a large complex of more than 20 proteins called the *mediator complex* is required to act as an intermediary (Figure 6.14) between the general transcription machinery (the RNA polymerase and the general transcription factors bound at the core promoter) and specific transcription factors bound to the proximal promoter or to more distant enhancers and silencers. It is assumed that the DNA bends and loops such that these transcription factors are brought into contact with the mediator complex.

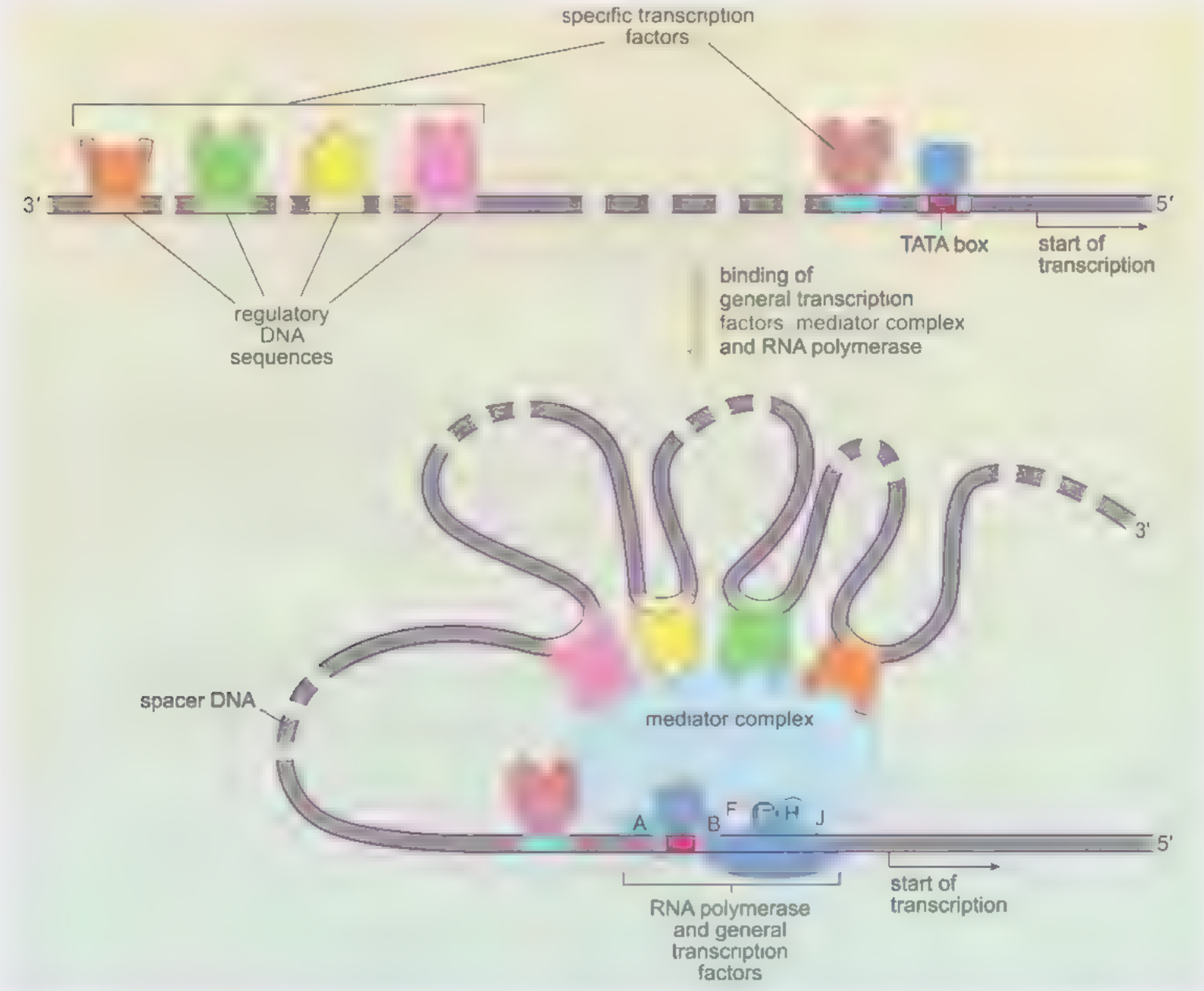



**Figure 6.14** Combinatorial control of eukaryotic gene expression by multiple transcription factors binding to regulatory elements. DNA bending at the promoter allows specific transcription factors bound to both local and distant regulatory DNA elements to interact with the RNA polymerase II complex, in many cases through binding to the mediator complex. Only the DNA template strand is shown, broken lines indicates long stretches of intervening DNA.

The whole complex of transcriptional activators, repressors and mediator complex is thus brought into contact with the RNA polymerase, and the different components of the complex interact to either stabilise or destabilise the assembly of the polymerase and general transcription factors at the core promoter. In addition, some transcriptional regulators also attract proteins that modulate chromatin structure to change the accessibility of the promoter to the general transcription factors and the polymerase, as will be discussed in the next section.

Eukaryotic transcription factors therefore largely work as an organised group to determine the expression of a gene in the appropriate cell, at the appropriate time and at the required level. This is often referred to as **combinatorial control**, and it has the advantage that many different genes can be controlled in complex ways with different combinations of a relatively small number of regulatory proteins. In many cases, however, it is useful to have a final key transcription factor that is able to switch a large number of genes on or off together in a coordinated fashion. An example of this is the glucocorticoid receptor, a protein that is expressed in almost all mammalian cells, but only binds to a specific regulatory DNA sequence when it forms a complex with a glucocorticoid hormone, for example cortisol which is released from the mammalian adrenal gland in response to stress. In the presence of cortisol, the glucocorticoid receptor coordinates the switching on of many different genes that increase the catabolism of sugars and fats to increase blood glucose, and also suppress the immune system; hence synthetic forms of cortisol are used to treat certain inflammatory diseases. You will learn more about the important role of combinatorial control and key transcriptional regulators in cell differentiation and development in Book 3, Chapter 1.

### 6.4.4 Accessibility of the [illegible] DNA to the transcription machinery

You will recall from Chapter 3 (Figure 3.12) that in eukaryotes, genomic DNA is packaged into chromosomes by wrapping the DNA around histone proteins to form structures called nucleosomes that are spaced along the DNA.

- What are the names given to transcriptionally inactive and active regions of chromatin structure?

  Heterochromatin is more densely packaged and condensed, generally contains few genes and is transcriptionally inactive, while euchromatin is more loosely packed and is where active gene transcription is taking place.

Wrapping of the DNA around histones imposes an obstacle to the binding of other proteins, such as the RNA polymerase and general transcription factors. In fact, some of the specific transcription factors activate promoters by changing the local structure of chromatin at the promoter to improve access to the DNA. This type of transcriptional activator binds to the histone proteins in nucleosomes and attracts enzymes called *histone acetylases* which catalyse the addition of acetyl groups ($CH_3CO$) to a particular type of amino acid (lysine) in the histone protein (there are also transcriptional repressors that attract

histone deacetylases, which remove these acetyl groups). The acetylation of histones is a signal that attracts *chromatin-remodelling proteins* whose function is to remove the nucleosomes or slide them apart to expose the DNA (Figure 6.15), allowing transcription factors and the RNA polymerase to bind to the DNA.

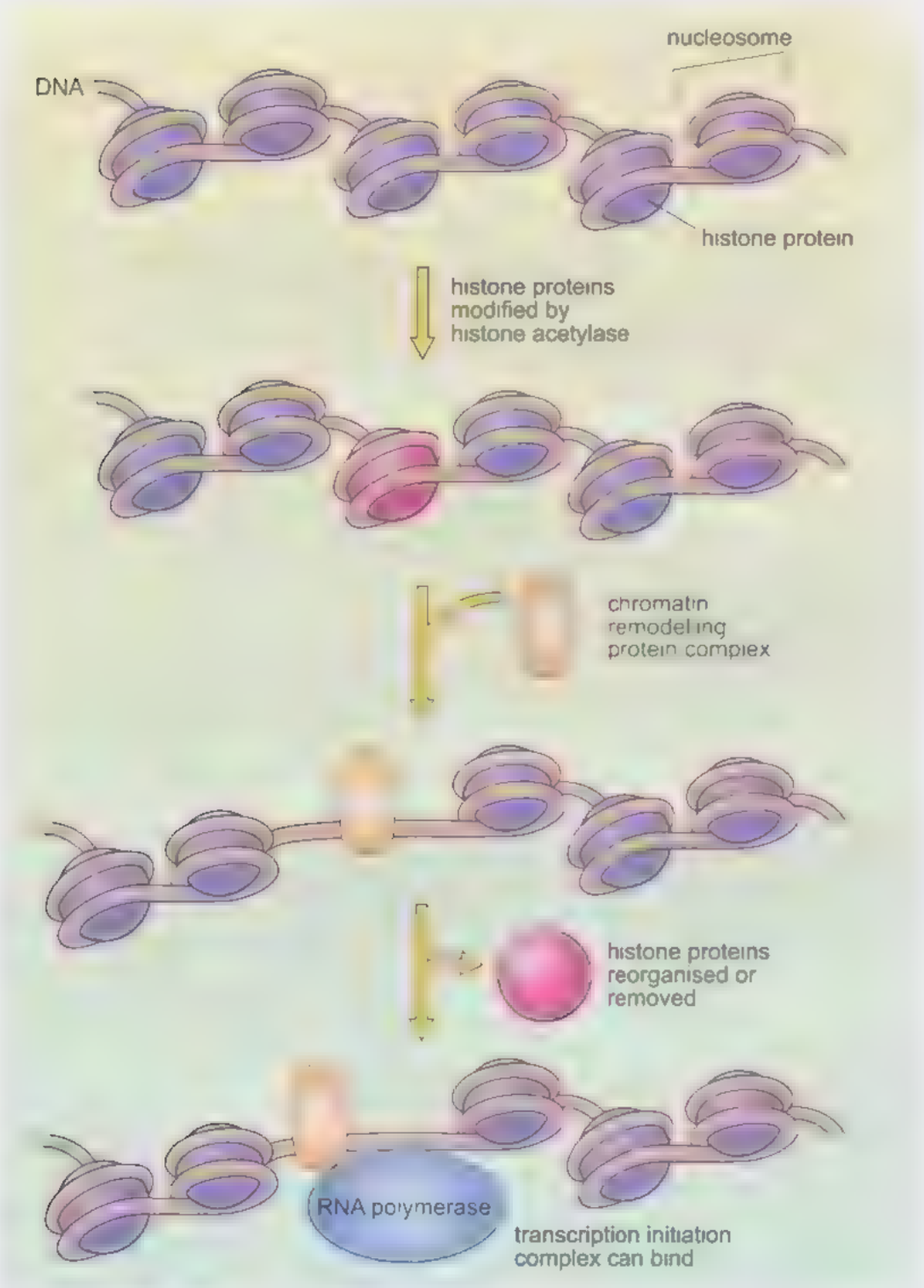



**Figure 6.15** Some transcriptional activators 'recruit' histone acetylases which modify histone proteins. The histone modification attracts chromatin-remodelling proteins to the promoter region which reorganise or remove histone proteins, making the DNA more accessible to the RNA polymerase.

## 6.4.5 The study of coordinate

Detecting the presence of the RNA transcript of a gene in a cell or tissue is the most direct way of determining whether the gene is active. To understand the role of individual genes in a particular type of cell or tissue, researchers have developed several methods for studying the level of mRNA of a single gene, or of multiple genes simultaneously (Box 6.2).

### Box 6.2 Methods for studying gene expression

#### Northern blotting

The British scientist Ed Southern invented a technique, called Southern blotting, in which DNA is separated by gel electrophoresis then transferred to a membrane. The corresponding technique used for RNA analysis was named 'northern' blotting in a play on words

The technique of **northern blotting** is used to separate and identify individual cellular mRNAs (Figure 6.16). First, all of the RNA is extracted from a sample of cells or tissue. Then the RNA is 'loaded' into a gel made of agarose, and an electric current is applied across the gel such that the negatively charged RNA molecules (negative because of the phosphate groups in their sugar backbone) are attracted towards the 'positive' electric pole. As the RNA molecules move through the gel towards the positive pole, the gel acts like a sieve, separating the RNA molecules on the basis of size (a technique called *gel electrophoresis*); small molecules move further through the gel than larger molecules.

The 'smear' of RNA molecules, separated according to size, is then transferred from the gel onto the surface of a nylon filter membrane. This membrane is then incubated with a solution containing a small DNA 'probe' sequence that is complementary to part of the mRNA of interest; a technique called *nucleic acid hybridisation*. The probe will interact only with the specific target mRNA sequence. The probe is first 'labelled' by incorporating either radioactive nucleotides, or nucleotides that are coupled to enzymes that can produce a coloured reaction product. If the mRNA of interest is present on the surface of the filter membrane, the labelled probe anneals with it by complementary base pairing, forming a DNA–RNA duplex. The location of the labelled probe can then be visualised using autoradiography (exposure to X-ray detecting film) in the case of a radioactive probe, or by the addition of a suitable chromogenic substrate for an enzyme-labelled probe. The main steps in northern blotting are shown in Figure 6.16.

**Figure 6.16** Northern blotting. (1) RNA is isolated. (2) The RNA molecules are separated on a size basis by agarose gel electrophoresis. RNA 'markers' are included so that the sizes of RNA species in the samples can be estimated. (3) The gel is compressed against a nylon filter membrane so that as the liquid from the gel soaks into the membrane by capillary action, the RNA moves out and is 'blotted' onto the membrane. (4) The blot is hybridised with a labelled probe to identify specific RNA molecules. Finally the filter membrane is processed to visualise the labelled probe attached to the target RNA molecule.


cells or tissue
1 extract RNA
RNA
2 agarose gel electrophoresis
marker RNA bands
sample RNA bands
3 blotting
weight
blotting paper
agarose ge
nylon filter membrane traps and binds nucleic acids
wick
buffer
support
buffer movement carries RNA from gel onto membrane
4 hybridisation and autoradiography
DNA probe hybridised to a specific RNA transcript

The advantages of the technique are that it identifies the size of the mRNA (by comparison with standard 'marker' RNAs of known size) and hence allows discrimination of alternatively spliced transcripts (Section 6.5.2). However, it can't be used to accurately quantify the amount of RNA and doesn't detect very low levels of mRNA.

### RT–PCR and qPCR

RT–PCR (reverse transcription polymerase chain reaction) has revolutionised the study of gene expression, making it possible to detect the RNA transcript of almost any gene, regardless of the scarcity of the starting material or relative abundance of the mRNA in the cell. It is a variant of the PCR technique described in Box 5.1. First, the total cellular mRNA is purified and used as a template for reverse transcriptase enzyme to generate DNA copies of the mRNAs called **complementary DNA (cDNA)**.

- What is a source of reverse transcriptase?

  This enzyme is produced by retroviruses which need it to produce a DNA copy of their RNA genome (Section 5.9).

The cDNA copy of the mRNA of interest can then be amplified using gene-specific primers in a PCR reaction (Figure 6.17). It is possible to accurately quantitate the amount of a particular mRNA by a quantitative version of PCR (qPCR). This is now routinely carried out using PCR primers labelled with a fluorescent 'tag'. Comparing the amount of fluorescent PCR product to a parallel 'standard' PCR reaction with a known starting amount of mRNA can produce an absolute measurement of the number of copies of the original target mRNA, typically in copies per cell. Some quantitative PCR techniques are referred to as 'real-time PCR' (often also rather confusingly abbreviated to RT–PCR) because the accumulation of the fluorescent PCR product can be monitored in real time.

The advantage of transcript detection by RT–PCR and qPCR is that it is quantitative and very sensitive (detection of a single mRNA molecule is possible), but it can be technically challenging, requiring careful design of experimental controls, because its extreme sensitivity means that even minute amounts of contamination by genomic DNA can produce misleading results.

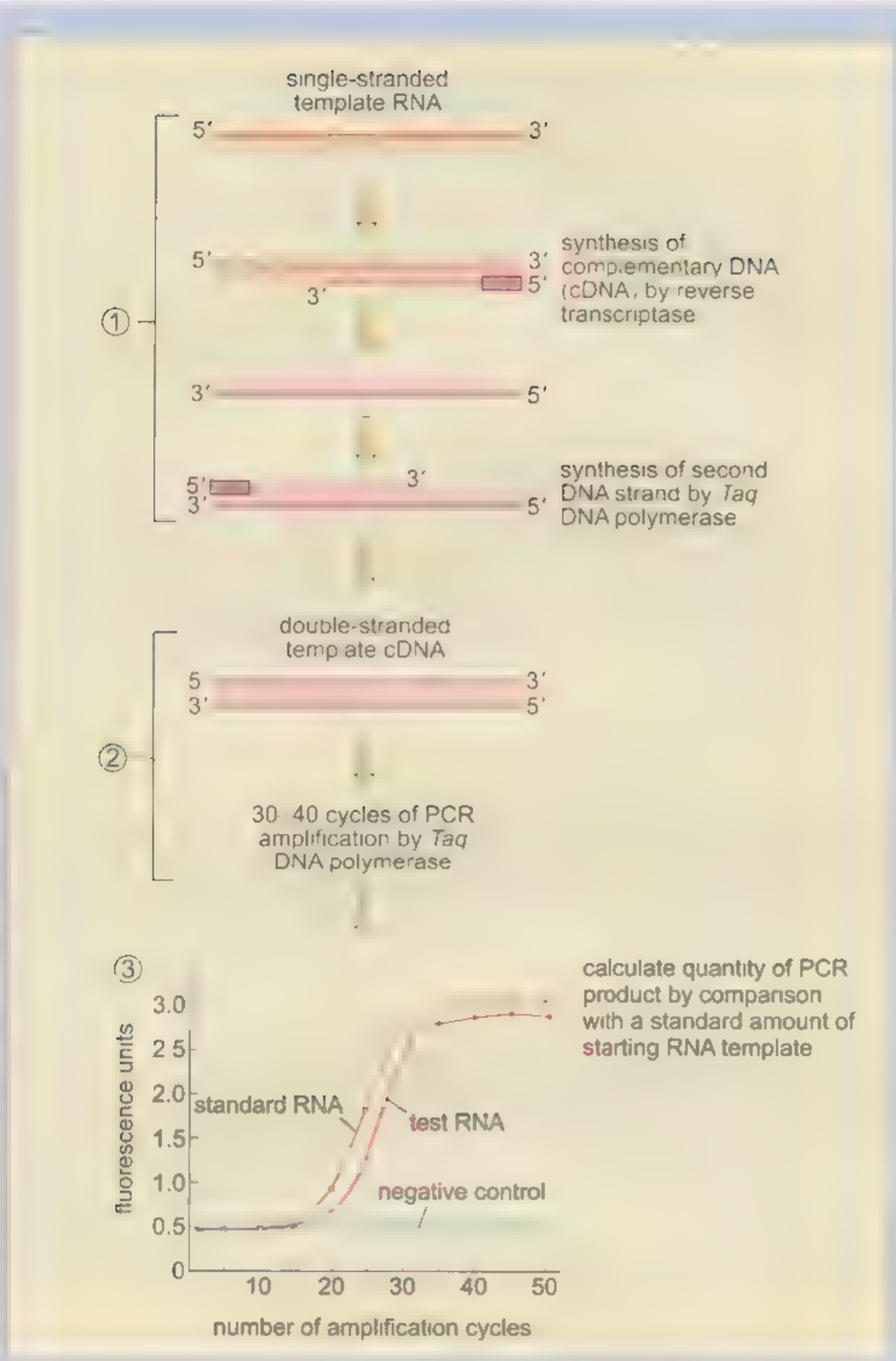



**Figure 6.17** RT–PCR and quantitative PCR. (1) The total RNA of the cells/tissue is isolated and reverse transcriptase enzyme synthesises a single-stranded complementary cDNA of every mRNA. *Taq* DNA polymerase then synthesises the second cDNA strand. (2) The double-stranded cDNA mixture is used as the template for multiple rounds of PCR using fluorescent-labelled primers specific for a particular gene sequence. (3) Comparison of the amount of fluorescent PCR product against a standard reaction starting with a known amount of mRNA allows the calculation of the amount of the specific mRNA in the original sample.

### DNA microarrays

The number of different genes that are activated in a single biological process may be huge. For example, during the process of transition from aerobic to anaerobic respiration (Book 2, Chapter 3) in the yeast *S. cerevisiae*, changes in the expression of 1740 genes have been recorded. The expression of multiple genes can be analysed simultaneously using a technique known as a **DNA microarray** or DNA 'chip'.

A DNA chip is a thin layer of silicon, 2 $cm^2$ or less in area, carrying a large number of individual DNA probes (usually more than 1000), each one with a different sequence representing a particular gene, and each at a defined position on the chip. The probes are synthetic single-stranded oligonucleotides that are spotted onto the silicon surface, using high-speed robots, to form the 'microarray'. The chip is incubated with cDNA prepared from total cellular mRNA using reverse transcriptase (see RT PCR above) and which has been labelled, usually with a fluorescent molecule. Hybridisation takes place between each probe and any of its target cDNA that is present in the sample. To determine how much target mRNA has hybridised to each probe spot, the surface of the chip is scanned, and the intensity of the fluorescent signal emitted by bound target mRNA is detected and recorded (Figure 6.18).

The DNA microarray analysis technique can be used to compare the pattern of mRNA expression in two different conditions – for example, the same cells treated in two different ways, cells from two different tissues, or diseased versus normal tissues (as illustrated in Figure 6.18). Studies on cancerous tissue using this technique discovered 299 genes with expression patterns that differed significantly when normal colon epithelial cells were compared with colon cancer cells. About half of these genes also showed abnormal expression levels in pancreatic cancer cells. The implication of this study is that some genes are abnormally expressed in more than one type of cancer, whereas others are abnormally expressed only in specific cancers. Because microarrays can be used to examine the expression of hundreds or thousands of genes at once, this technique has revolutionised the way scientists examine gene expression, and identify genes involved in diseases.

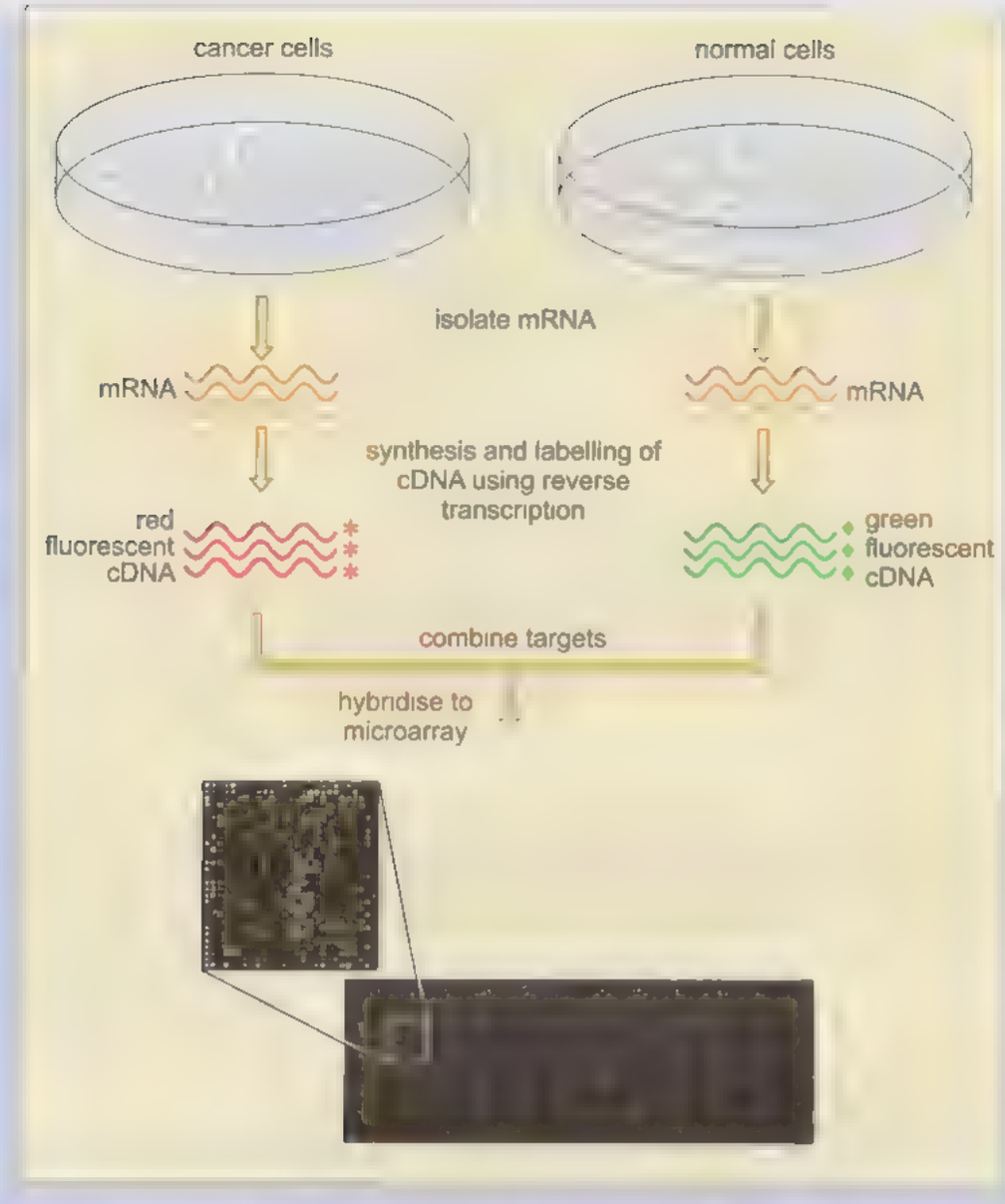



**Figure 6.18** DNA microarrays allow the expression level of multiple genes to be determined in a single assay. cDNA is prepared by reverse transcription of template mRNA purified from cells or tissues, labelled with a fluorescent molecule and hybridised to the microarray chip. Labelled mRNAs bound to the individual DNA probes on the chip are detected by fluorescence microscopy, and the intensity of the fluorescence signal measured. This example shows two-colour detection to compare gene expression in normal and cancer cells. The cDNAs prepared from the two cell types have been labelled with different fluorescent probes (red and green) so that differences in gene expression between the two cell types are detected by the relative intensity of red and green fluorescence associated with each DNA probe spot on the chip.

## and termination

Once initiation is complete, the process of elongating the new RNA strand is similar in both prokaryotes and eukaryotes (see Section 6.3.1). Elongation is not a smooth process. The polymerase transcribes rapidly through some sequences but pauses at others.

■ What features of eukaryotic DNA might impede the RNA polymerase?

Regions where the DNA is wrapped around nucleosomes.

During elongation in eukaryotic cells, the RNA polymerase is associated with *elongation factor* proteins that reorganise the nucleosomes. This allows the polymerase to progress along the DNA and ensures that it doesn't dissociate from the DNA before it reaches the end of the transcription unit.

Considerably less is known about the control of transcription termination in eukaryotes than in bacteria. The three RNA polymerases use different mechanisms. Termination of the transcription of mRNAs by RNA polymerase II is coupled to the process that polyadenylates the 3′ end of the transcript (Section 6.5.1 below).

There are also some examples in eukaryotes where protein levels are regulated by a mechanism that results in premature termination of transcription. An example of this occurs during transcription of the HIV-1 (AIDS virus) genome in the eukaryote cells that it infects. During early infection, when the disease is not fully developed, the viral mRNA molecules terminate prematurely, and are not translated into functional proteins, so few new virus particles are made. However, in the later stages of infection, high levels of full-length transcripts are produced because premature termination is overridden by an HIV-1 protein called Tat, a mechanism referred to as *anti-termination*.

## Summary of Section 6.4

- There are three forms of the enzyme RNA polymerase in eukaryotes, namely RNA polymerase I, II and III. RNA polymerase II transcribes protein-coding genes.
- All three eukaryotic RNA polymerases contain three types of core subunit (with sequence homology to prokaryotic RNA polymerase subunits) and a number of smaller subunits.
- The core promoters of most eukaryotic protein-coding genes have a highly conserved sequence called the TATA box (homologous to the prokaryotic −10 box) positioned between 25 and 35 bp upstream of the transcription start site.
- Formation of the transcription initiation complex requires sequential assembly of general transcription factors at the core promoter, starting with TFIID and culminating in the binding of RNA polymerase II to the core promoter.
- Changes in the structure of chromatin to remove the obstacle presented by nucleosomes is necessary for proteins involved in transcription to gain

access to eukaryotic DNA, and chromatin modification contributes to the regulation of gene expression.

- Specific transcription factors bind to constitutive and regulatory DNA elements in the upstream promoter, and to enhancer and silencer sequences which may be even greater distances from the promoter. These regulate transcriptional initiation either by stabilising or destabilising binding of the RNA polymerase and general transcription factors to the core promoter, or by attracting proteins that modify chromatin locally to make it either more or less accessible.
- The level of gene expression in eukaryotes (as in prokaryotes) is largely determined by the regulation of transcriptional initiation, but in some cases can also be controlled by elongation factors or termination mechanisms.

## 6.5 Post-transcriptional control of gene expression in eukaryotes

Prokaryotic mRNAs are ready for translation without further processing. Indeed, prokaryotic mRNAs start to be translated even before transcription is completed.

- Bearing in mind the differences between prokaryotic and eukaryotic intracellular structure, why is coupling of transcription and translation possible in prokaryotic cells but not in eukaryotic cells?

- Prokaryotes have no nuclear envelope separating their DNA from the cytoplasm. In contrast, transcription takes place in the nucleus of eukaryotic cells, so the mature mRNAs must be transported out of the nucleus to the cytoplasm before translation can take place at ribosomes in the cytoplasm.

Transcription is really only the first step in the production of a mature eukaryotic mRNA. The primary transcript must undergo a number of modifications including removal of intron sequences, *capping* of the 5′ end and *polyadenylation* of the 3′ end (Figure 6.5) before it is ready for export from the nucleus to the cytoplasmic ribosomes. These processing events and their regulation are described briefly in this section.

### 6.5.1 Modification of the ends of mRNA

**Capping** is a process that occurs immediately after transcription has started. Eukaryotic mRNA is capped at its 5′ end by the addition of an unusual modified guanosine triphosphate (GTP) nucleotide. The reaction is catalysed by the enzyme guanylyl transferase, which is a component of the RNA polymerase II complex. The 5′ cap has several important functions in eukaryotes:

- It is required for mRNA export through the nuclear pores.
- It is essential for ribosome binding and efficient translation of the mRNA (Section 6.6).

- It prevents enzymes in the cytoplasm from degrading the mRNA from its 5′ end before it can be translated.

Capping can therefore also regulate protein expression by regulating mRNA half-life and translation. An interesting example of how the cap can be used in gene regulation comes from studies on influenza virus, which contains RNA as its genetic material. This virus does not cap its own genome segments, but steals pre-formed caps from host mRNA, a process called 'cap-snatching'. The host mRNA, which has lost its cap, cannot bind to ribosomes and initiate translation (Section 6.6.1). Thus, host protein can no longer be made, but viral protein can.

Almost all eukaryotic mRNAs (and some non-protein-coding RNAs) also undergo **polyadenylation** of their 3′ end, immediately after termination of transcription and release of the primary transcript. Polyadenylation involves addition of multiple adenosine nucleotides to generate a **poly(A) tail**. The length of the tail may vary; but around 200 adenosines is usual. The poly(A) tail acts as the binding site for poly(A) binding protein which inhibits degradation of mRNA from the 3′ end and can also promote its export from the nucleus, and its translation.

## 6.5.2 mRNA splicing

You will recall from Section 5.8.1 that in eukaryotic genes, the sequences encoding the polypeptide (exons) are interrupted by non-coding introns. The introns must therefore be removed from the primary transcript to produce a mature mRNA suitable as a template for translation. Introns are removed by mRNA splicing, a process in which the introns are cut out and the exons are joined together to form the mature mRNA. Splicing is carried out by a complex known as the *spliceosome*, containing several proteins and small nuclear RNAs (snRNAs, Section 6.2.1). The snRNAs bind to complementary sequences at the ends of introns, and the spliceosome then cuts out each intron and joins the ends of the flanking exons together.

In fact, mRNA splicing also has an important role in gene regulation, because the outcome of splicing may produce a number of different transcripts depending on which exons are included in the mature transcript (Figure 6.19). This differential or *alternative splicing* may therefore result in the expression of several different mRNAs from the same gene in different cells and tissues. The mRNAs produced in this way (which share some exons and not others) are called *splice variants* or *splice isoforms*. The resulting variant mRNAs may be translated into different protein isoforms; thus, a single gene may code for multiple proteins which may have similar or different properties.

Gene control by tissue-specific alternative RNA splicing has been observed in many eukaryotic genes, the fibronectin gene is a good example. Fibronectin is an extracellular matrix protein secreted by fibroblasts and liver cells (as well as other types of cells). The form of fibronectin secreted by fibroblasts can interact with cell surface receptors called integrins (Section 3.5.1) and helps cells to adhere to the extracellular matrix. The form of fibronectin that is secreted by liver cells, in contrast, does not promote adhesion, but is secreted into the circulation and is a major component of blood plasma. Both types of

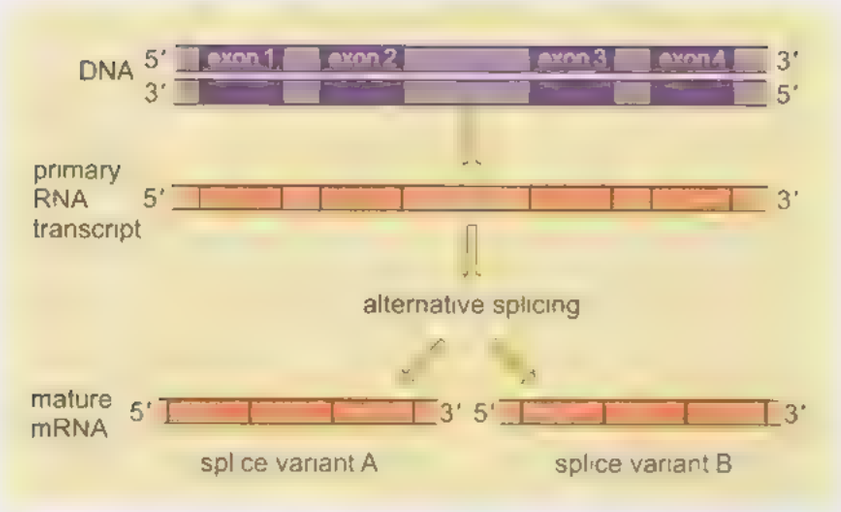



**Figure 6.19** Alternative splicing of eukaryotic mRNA. The gene in this example contains four protein-coding exons (dark purple bands) separated by non-coding introns (light purple bands). The same sequence is transcribed into the primary mRNA, which then undergoes splicing to yield mature mRNAs. Two splice-variant mRNAs are shown that have resulted from splicing the same primary transcript in different ways to retain different combinations of exons.

protein are encoded by the same fibronectin gene, but the primary transcript is subject to alternative splicing. In fibroblasts, splicing yields a mature mRNA that includes two exons encoding the protein domains that interact with cell surface receptors in several cell types. Splicing of the fibronectin primary transcript in liver cells, however, 'skips' these two exons, yielding a mature mRNA that does not encode the corresponding domains.

## 6.5.3 The stability of mRNA

At one time, biologists generally assumed that the rate at which a particular mRNA was synthesised was the major determinant of the amount of that mRNA in the cell, and therefore the amount of the corresponding protein in the cell. Further research has, however, demonstrated that the rate of mRNA *turnover* can also significantly affect the rate of protein synthesis. In other words, the production of proteins reflects not only how fast their mRNA templates are made, but also how fast the mRNA is broken down, or degraded. In eukaryotic cells there is a balance between the processes of translation and mRNA degradation.

- How are mRNAs that are being actively translated protected from degradation?

  By binding to ribosomes and poly(A) binding protein and by 5′ capping (Section 6.5.1).

Messenger RNAs are constantly attacked by enzymes called **ribonucleases** that remove nucleotides from the end of the poly(A) tail and gradually shorten it over time, so that the mRNA eventually becomes vulnerable to complete degradation.

Regulated changes in the turnover rates of various mRNAs can also help the cell to respond quickly to changes in its environment. An extreme example of this is the herpes simplex virus which causes cold sores and some genital infections. Like other viruses, it uses the molecular machinery of the infected host cell to replicate itself. It appropriates the host cell's protein synthesis machinery and eliminates competing demands by producing a viral ribonuclease that preferentially degrades the host cell's mRNA molecules, thereby freeing up more ribosomes to translate viral mRNAs.

In some cases, a protein may regulate the level of its own mRNA. An example of this is tubulin. As you learnt in Section 3.4.2, tubulin monomers assemble into microtubules, which are part of the cell's cytoskeleton. When a large pool of free tubulin monomers is available, there is no need for the cell to make more. Under these conditions, tubulin molecules bind to tubulin mRNA, making it susceptible to breakdown by ribonucleases, and reducing the synthesis of new tubulin protein.

### 6.3.4

It has become increasingly apparent in recent years that certain types of non-coding RNA (Section 6.2.1) are important regulators of gene expression. The most extensively studied of these are **microRNAs (miRNAs)**. These are very small ncRNAs, about 22 nucleotides long, found in the cells of animals and plants, which can negatively regulate, or 'silence' the expression of certain genes.

MicroRNAs are themselves encoded by genes that often occur in clusters. These are transcribed as long, polycistronic primary transcripts that are cleaved; first in the nucleus (by a ribonuclease called Drosha), and then in the cytosol (by another ribonuclease called Dicer) to finally generate ~22 base pair double-stranded miRNA molecules (Figure 6.20). One strand of each miRNA is then incorporated into a protein complex called *RNA-induced silencing complex (RISC)* and guided to bind to 3′ regions of complementary target mRNAs, resulting in either degradation of the target mRNA or inhibition of its translation. miRNAs therefore provide yet another mechanism of regulating gene expression. Regulation by miRNAs is not as simple as one miRNA per gene. Each miRNA can potentially target multiple mRNAs. There are more than 700 miRNAs encoded in the human genome that could, in theory, collectively regulate at least a third of human genes.

Although the function of most miRNAs is yet to be determined, many seem to be important during the development of different tissues in multicellular organisms. Different miRNAs may coordinate the regulation of a gene at different times during development. One particular miRNA, known as miR-1, is expressed in developing heart muscle and regulates the level of several muscle-specific transcription factors that control muscle protein expression. In a strain of mutant mice that do not express miR-1, the cardiac muscle precursor cells, i.e. the cells destined to differentiate into specialised cardiac muscle cells, proliferate a great deal more than normal, but are unable to differentiate into mature functional muscle. This results in severe heart defects. miR-1 therefore seems to 'fine-tune' the balance between the

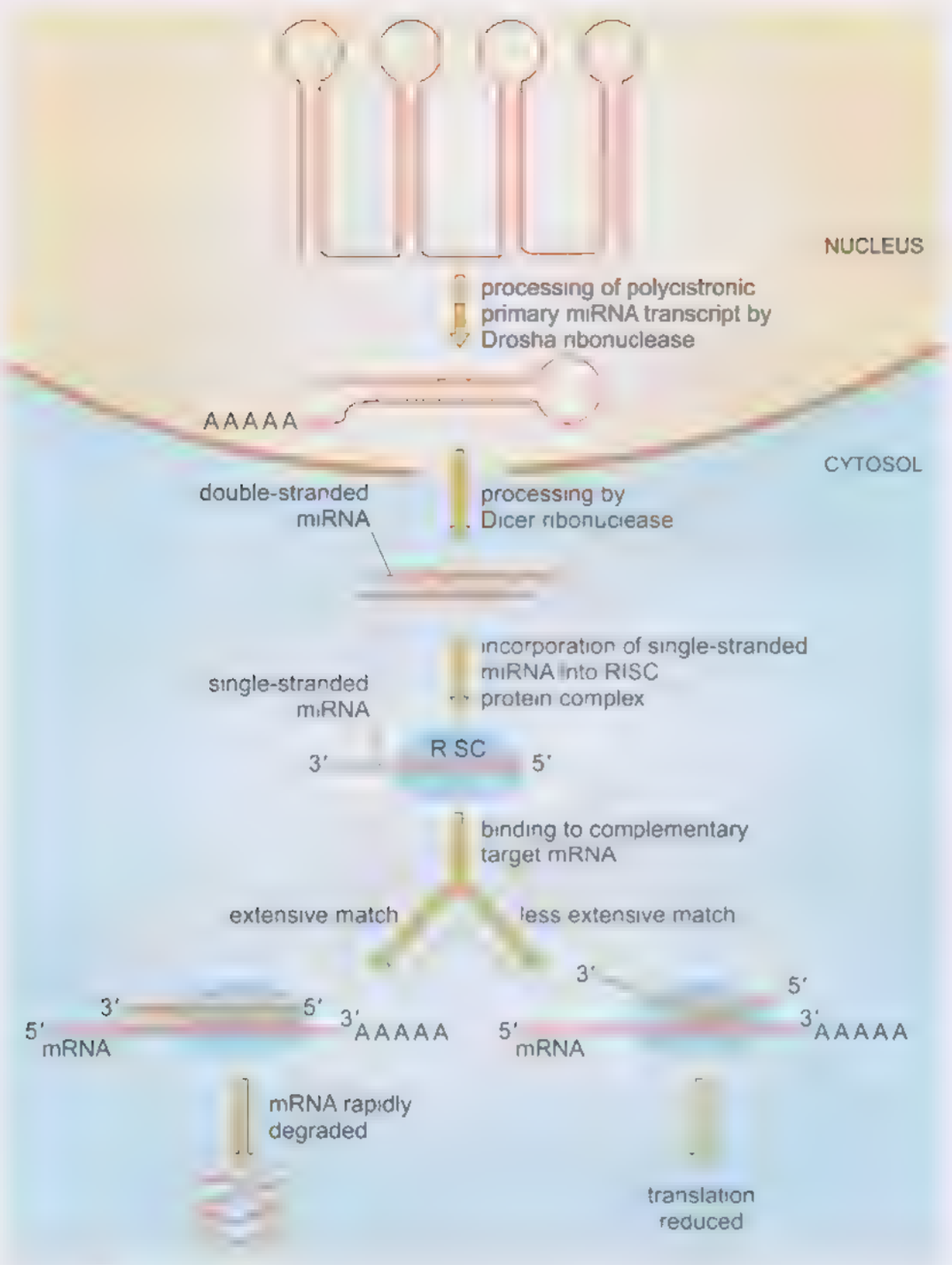



**Figure 6.20** The expression and processing of miRNAs. A polycistronic primary transcript including several miRNA sequences is processed in the nucleus by the Drosha ribonuclease to separate the hairpin-shaped pre-miRNAs. These are processed further in the cytoplasm by the Dicer ribonuclease to form ~22 base pair double-stranded miRNAs which associate with the RISC complex as single strands. The RISC complex then recognises mRNAs complementary to the miRNA strand and, depending on the degree of complementarity between the miRNA and the target mRNA, either blocks their translation, or degrades the mRNA.

proliferation of cardiac muscle precursors and their differentiation into mature muscle cells. You will learn more about cell differentiation in Book 3, Chapter 1.

There is evidence that miRNAs act as key regulators of a wide range of processes, and not surprisingly, they have been implicated in a number of diseases including heart disease, neurological diseases and cancers. The ability of complementary RNAs to regulate gene expression has also been exploited for research into gene function (Box 6.3).

## Box 6.3 Technologies using antisense RNA, miRNA and siRNA

The ability of RNA molecules to specifically switch off the expression of individual genes has exciting potential for cell biology research. The first technique of this type to be developed was the use of **antisense RNA**, in which a single-stranded RNA complementary to a specific mRNA is introduced into cells. The double-stranded hybrid RNA that forms cannot be recognised by the ribosome, so mRNA translation and synthesis of a polypeptide are blocked. However, this technique has proved difficult to develop for commercial and clinical applications.

While researchers were conducting experiments to test the effects of antisense RNAs, it became apparent that the complementary sense RNA (used in control experiments) unexpectedly turned out to be just as effective as the antisense RNA. The reason is probably that these experiments were contaminated by double-stranded RNAs (dsRNAs) formed from sense and antisense duplexes. The ability of dsRNA to specifically suppress the expression of a gene containing the same sequence is called **RNA interference (RNAi)**. When the dsRNA enters a cell, it is cleaved by Dicer into fragments about 22 bp long called small interfering RNAs (siRNAs). The two strands of the siRNA separate, and the antisense strand is incorporated into the RISC complex. The RISC complex then degrades any mRNAs containing sequence complementary to the siRNA.

- In what ways is the processing of siRNAs and miRNAs similar?

- Both are processed by Dicer in the cytoplasm to produce ~22 bp duplex one strand of which is incorporated into the RISC complex.

Small synthetic double-stranded siRNA molecules can be easily synthesised in the laboratory and the RNAi technique is routinely used to turn off expression of individual genes in order to explore their function in cells or whole organisms (Box 4.1). siRNA 'libraries' are available which include siRNAs capable of targeting almost every individual gene in the genomes of *C. elegans*, *Drosophila* and the human genome.

miRNAs and siRNAs also have potential for gene therapy. Research has identified siRNAs that could be used to treat the eye condition acute macular degeneration (AMD). siRNAs that repress the expression of the growth factor VEGF prevent the overgrowth of blood vessels in the retina of the eye that can result in vision loss in people with AMD. The expression of some other natural miRNAs is defective in certain cancers, and artificial expression of a specific miRNA in mice with an aggressive

form of liver cancer was shown to halt the disease. However, there are still many practical problems to solve before small RNAs become useful for treating disease, including how to efficiently deliver artificial miRNA or siRNA into the body, and how to avoid unexpected effects on other genes.

## Summary of Section 6.5

- Most eukaryotic mRNAs are first produced as primary transcripts which must be processed to produce a mature RNA. This is achieved by capping (addition of a GTP residue at the 5′ end), polyadenylation (addition of a variable number of adenosine nucleotides to generate a poly(A) tail at the 3′ end) and splicing to remove the non-coding introns from the mRNA.
- Availability of mRNA is an important factor in gene regulation; the production of proteins reflects not only how fast their mRNA templates are made, but also how fast they are degraded.
- Degradation of mRNA is a specific process that removes coding mRNAs whose protein products are no longer required. mRNA stability (and therefore half-life) is affected by modifications like capping and polyadenylation.
- MicroRNAs (miRNAs) have an important role in post-transcriptional regulation of gene expression, by targeting specific mRNAs for degradation, or by inhibiting their translation. This phenomenon can be exploited experimentally by introducing chemically synthesised small interfering RNAs (siRNAs) into cells in order to target specific mRNAs. This allows researchers to turn off the expression of specific genes to study the effects on a cell or organism.

## 6.6 Translation

Translation is the process by which the genetic information carried in an mRNA is translated into a polypeptide chain. Each amino acid in the polypeptide is encoded by a triplet codon.

- How many different codons are there?

  64. There are four different types of nucleotides (A, G, C and U/T) which can exist in $4^3$ different combinations of three.

As described previously (Section 6.2.2), transfer RNAs (tRNAs) form the link between the mRNA sequence and the polypeptide sequence. The terminology used to describe specific tRNAs uses a superscript to denote the amino acid that is carried by the tRNA. For example, a tRNA carrying methionine is denoted $tRNA^{Met}$.

The ribosome has a similar structure in both prokaryotes and eukaryotes, and consists of a large complex of ribosomal RNAs and proteins (more than 50 proteins in prokaryotes and more than 80 proteins in eukaryotes) organised

into a large and a small subunit (Figure 3.15). As with gene transcription, translation can be divided into three stages: initiation, elongation and termination of the polypeptide chain. The last two stages are similar in prokaryotes and eukaryotes, but initiation is somewhat different, as will be described in the next two sections.

## 6.6.1 Initiation of translation in prokaryotes

The individual subunits of ribosomes are present free in the cytoplasm, and for translation to be initiated, a complete ribosome must assemble at the appropriate place on an mRNA molecule. Attachment of the ribosome to the mRNA occurs at the **ribosome binding site**, which in *E. coli* has the consensus sequence 5′ AGGAGGU 3′ (Figure 6.21a). This site lies 3–10 nucleotides upstream of the translation initiation codon – the first codon in the protein-coding open reading frame of the mRNA. The initiation codon is almost always 5′ AUG 3′, which codes for the amino acid methionine.

In prokaryotes, the small (30S) subunit of a ribosome binds first to the ribosome binding site on the mRNA, with the help of two *translation initiation factor* proteins called IF1 and IF3 (Figure 6.21b) which also prevent the large ribosomal subunit from binding too soon. Next, the initiator tRNA carrying methionine ($tRNA_i^{Met}$) is brought to the small subunit of the ribosome by another translation initiation factor, IF2, which is bound to a molecule of GTP (which is an energy-rich molecule similar to ATP, Section 1.2.1) (Figure 6.21c). The large (50S) subunit of the ribosome then attaches to the small subunit using the energy stored in the GTP molecule bound to IF2 (Figure 6.21d). During this process, the GTP is hydrolysed to GDP by breaking the bond to one of its phosphate groups (releasing the energy stored in the bond). The IF1, IF2 and IF3 dissociate from the ribosome complex and translation initiation is complete.

## 6.6.2 Initiation of translation in eukaryotes

In contrast to prokaryotic mRNA, eukaryotic mRNA does not have a ribosome binding site, so the initiation codon has to be located by another mechanism. In eukaryotes, a complex including the small (40S) ribosomal subunit, an initiator $tRNA_i^{Met}$, the eukaryotic translation initiation factor eIF2 (bound to GTP) is formed independently in the cytoplasm (Figure 6.22a). Initiation of translation usually involves the attachment of this complex to the mRNA at the 5′ cap structure, a process requiring two more translation initiation factors called eIF3 and eIF4 (Figure 6.22b). The whole complex then scans along the mRNA until it locates the first translation initiation codon, 5′ AUG 3′. Selection of the correct 5′ AUG 3′ is facilitated by recognition of the consensus sequence, 5′ ACC*AUG*G 3′ containing the initiation codon (italicised). This consensus sequence is called the Kozak sequence after Marilyn Kozak, the scientist who identified it.

In some cases, translation initiation is cap-independent, and depends instead on internal sequences in the RNA, but this mechanism will not be described in detail here.

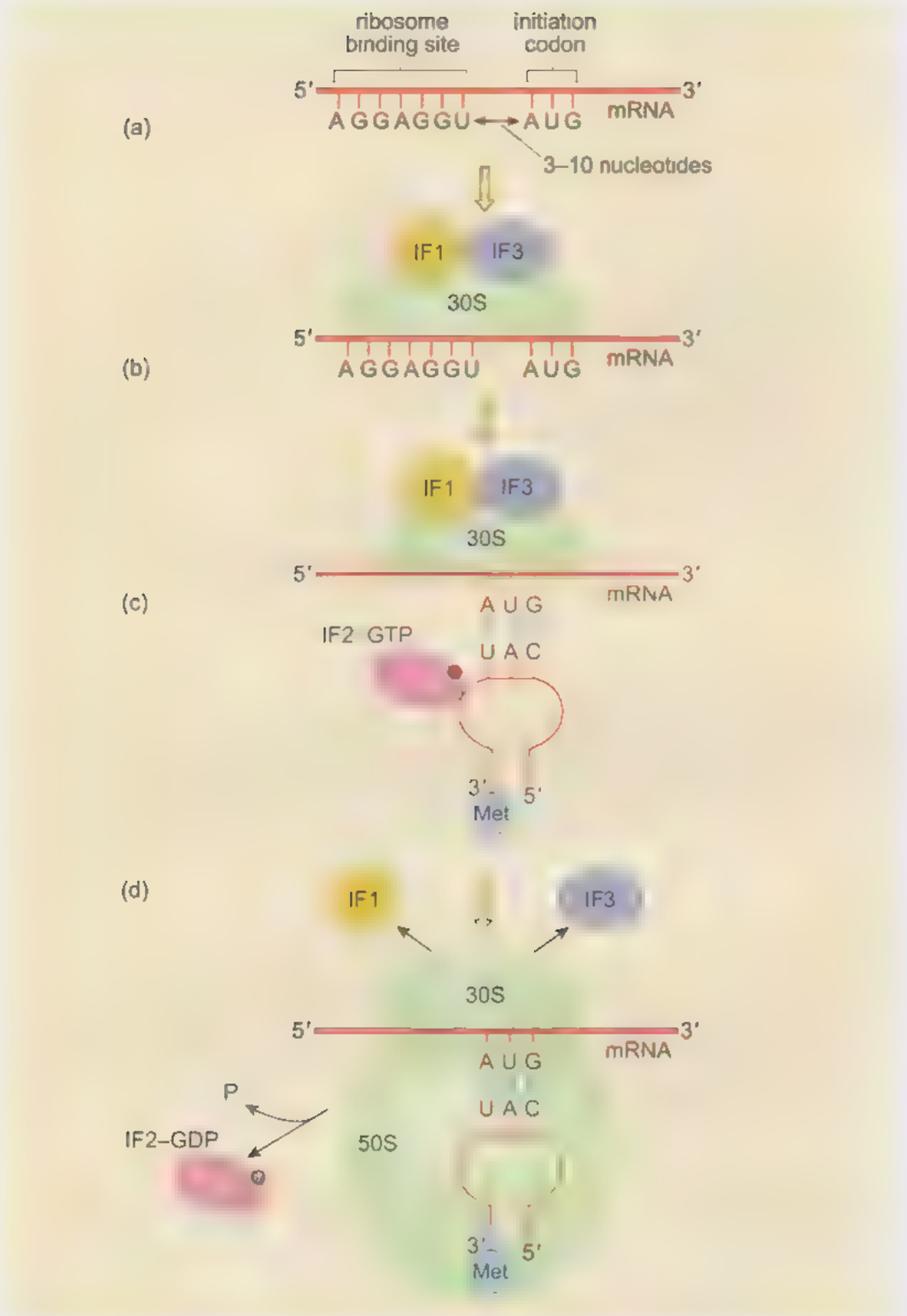



**Figure 6.21** Initiation of translation in prokaryotes (a) and (b) The small (30S) ribosomal subunit and the translation initiation factors IF1 and IF3 bind to the ribosome binding site (c) The initiator tRNA is then brought in by IF2 and binds to the start codon (AUG). (d) The GTP molecule (red dot) bound to IF2 is hydrolysed to GDP (grey dot) and inorganic phosphate to provide energy for binding of the large (50S) ribosomal subunit, and the translation initiation factors dissociate from the complex, which completes initiation.

Once the initiation codon is recognised the GTP bound to eIF2 is hydrolysed to provide energy for the attachment of the large (60S) ribosomal subunit, and at the same time the eIF2, eIF3 and eIF4 factors all dissociate from the small ribosomal subunit (Figure 6.22c). Initiation is now complete and the ribosome commences polypeptide elongation.

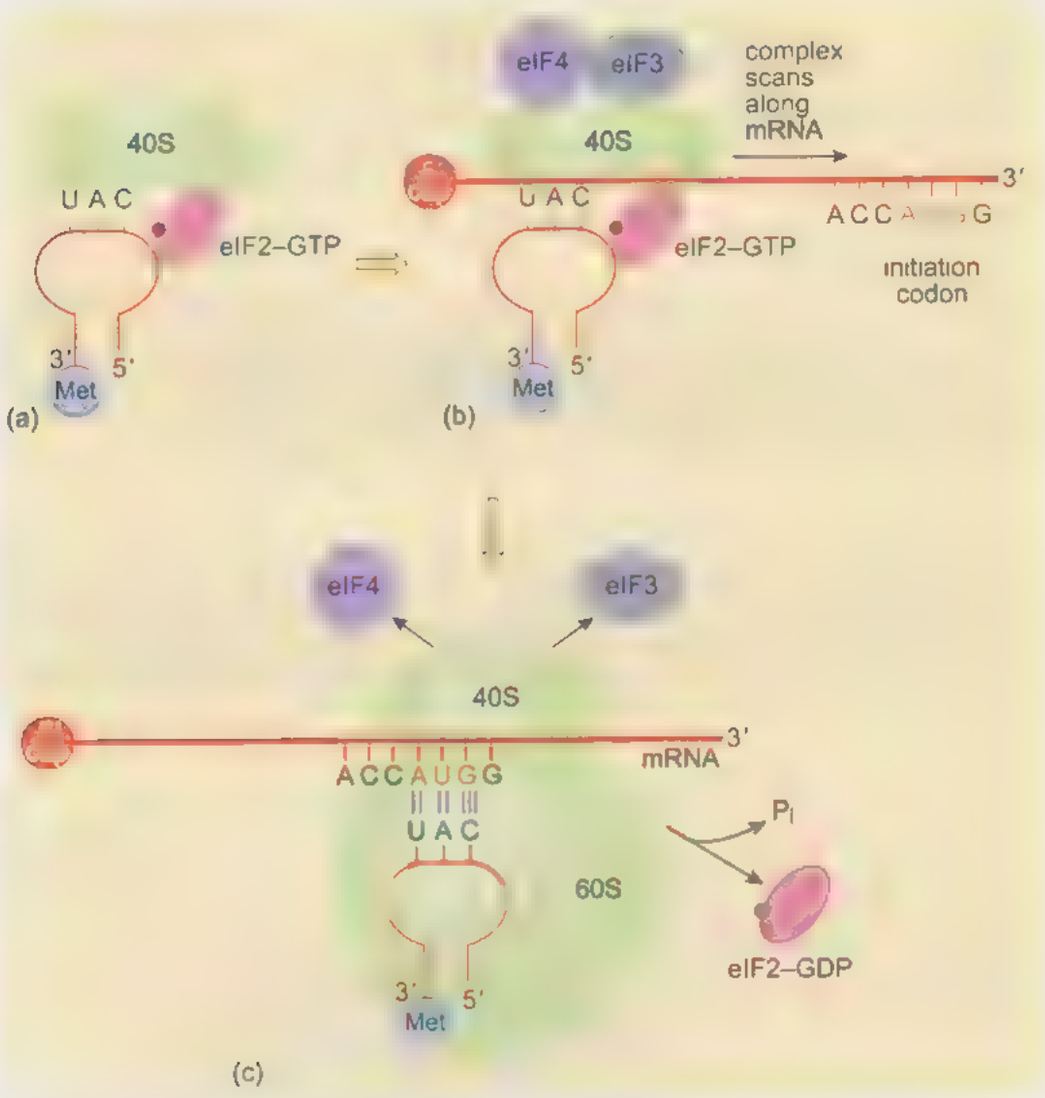



**Figure 6.22** Initiation of translation in eukaryotes. (a) and (b) A preformed complex comprising the small (40S) ribosomal subunit, the translation initiation factor eIF2 and the initiator tRNA attaches to the 5′ cap structure of the mRNA (facilitated by eIF3 and eIF4) and scans along to locate the start codon (AUG). (c) The GTP molecule (red dot) bound to eIF2 is hydrolysed to GDP (grey dot) and inorganic phosphate to provide energy for binding of the large (60S) ribosomal subunit, and the translation initiation factors dissociate from the complex which completes initiation.

## 6.6.3 Elongation of the polypeptide

Polypeptide chain elongation is similar in prokaryotes and eukaryotes and involves a great many protein factors, so the full details are not explored here. The large ribosomal subunit attached to the mRNA has two binding sites for

tRNAs: a P (polypeptide) site and an A (acceptor) site. The P site is shown already occupied by $tRNA_i^{Met}$ in Figure 6.23a. A translational elongation factor (called EF-Tu in prokaryotes and eEF1 in eukaryotes) bound to a GTP molecule brings in a second tRNA molecule with the appropriate anticodon to the A site (Figure 6.23b). If the codon–anticodon pairing is correct, the elongation factor hydrolyses GTP to GDP and inorganic phosphate, creating a conformational change in the ribosome that causes the incoming tRNA to fully enter the A site. The ribosome then catalyses the formation of a peptide bond between the methionine at the P site, and the second amino acid at the A site (phenylalanine in the example shown in Figure 6.23c).

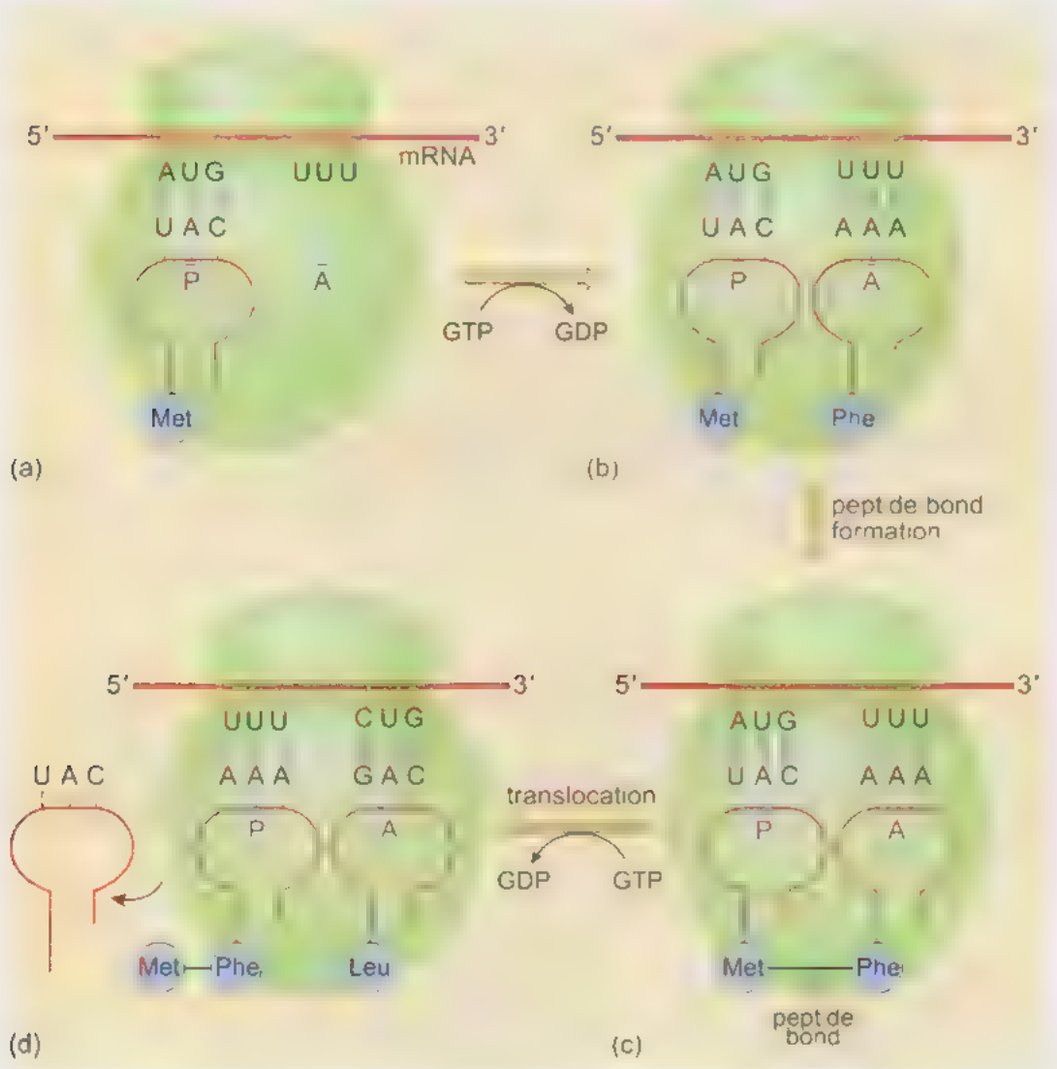



**Figure 6.23** Elongation of translation (see description in the text).

The next step is *translocation*, during which three things happen (Figure 6.23d):

i The mRNA moves along three nucleotides so that the $tRNA^{Phe}$ originally in the A site moves to the P site, and the next codon in the mRNA (CUG in the example shown) now enters the A site and a new tRNA ($tRNA^{Leu}$) occupies it.

ii The amino acid bound to $tRNA_i^{Met}$ is released.

iii The $tRNA_i^{Met}$ leaves the ribosome.

The translocation step is facilitated by another type of elongation factor (called EF-G in prokaryotes and eEF2 in eukaryotes) which also carries a GTP molecule. The elongation factor hydrolyses GTP causing a conformational change in the ribosome that triggers translocation. A new peptide bond is formed with the next amino acid. Repetition of this cycle elongates the polypeptide chain and synthesis continues towards the carboxy terminus (C-terminus) until a termination codon is reached.

Messsenger RNAs that are translated at free ribosomes in the cytoplasm are mostly found in **polyribosomes**, often abbreviated to **polysomes**. These are clusters of ribosomes that are all bound to a single mRNA molecule. As soon as one ribosome has moved far enough away from ribosome binding site, another ribosome is able to bind and initiate translation, so that many ribosomes can be spaced along the mRNA, each synthesising a polypeptide chain (Figure 3.16). This multiple initiation from one transcript means that new protein molecules can be produced very quickly.

- In eukaryotes, proteins destined for lysosomes, or for the cell membrane or for export from the cell are synthesised on ribosomes attached to the rough endoplasmic reticulum (RER). How are these proteins directed to the RER?

  These proteins contain a RER localisation signal sequence in the amino terminus of the translated protein, so while their translation is initiated on free ribosomes in the cytoplasm, as soon as the signal sequence is recognised (by the signal recognition particle) the whole complex of ribosome, mRNA and partially translated polypeptide is relocated to the RER to complete translation.

### 6.6.4 Termination of translation

When a termination codon is reached, a *release factor* (RF) protein enters the A site of the ribosome instead of a tRNA. The RF recognises one of the three termination or stop codons, UAA, UAG or UGA, and ends elongation, releasing the completed polypeptide from the tRNA in the P site. The ribosomal subunits are released back into the cytoplasmic pool.

### 6.6.5 [illegible] gene expression

As you have learnt, mechanisms that control the amount of mRNA produced by transcription are crucial for controlling gene expression. If, however, a particular mRNA is present but there is no need for the cell to make the protein, it is wasteful and may indeed be harmful for the mRNA to be translated. There are several mechanisms that regulate the initiation of translation of some mRNAs. For example, the recruitment of the small ribosomal subunit can be inhibited by secondary structure in the mRNA (loops and hairpins formed by complementary base pairing), or by specific RNA binding proteins. The control of the production of the iron storage protein ferritin in certain mammalian cells is a good example of the latter mechanism.

If the iron level in the cell is low, ferritin is not required and a translational repressor protein binds to the ferritin mRNA and prevents its translation by blocking its attachment to the ribosome. When iron levels rise, the free iron binds to the repressor and alters its conformation, causing it to detach from the mRNA. Translation of ferritin protein then proceeds.

Translated proteins that are incomplete or misfolded may be disruptive or even toxic and are rapidly degraded in the cell by a very large protein complex known as the **proteasome**. Most proteins are eventually degraded by the proteasome when they are no longer needed. The half-life of a protein largely depends on its N-terminal amino acid, and whether it possesses certain sequences. For example, regions rich in the amino acids Pro (P), Glu (E), Ser (S) and Thr (T) (called PEST sequences) mark the protein for rapid degradation. The balance between protein synthesis and protein degradation (protein turnover) contributes to the regulation of protein levels and ensures that old, damaged proteins don't accumulate in the cell.

Translated polypeptides, particularly in eukaryotes, undergo a whole range of regulated post-translational modifications. Many polypeptides are not active when they are first formed but require additional folding, binding to other polypeptides, enzymatic cleavage, or modifications, such as phosphorylation or glycosylation of particular amino acids, before they become fully active. There is not space to discuss post-translational protein modification and protein turnover here, but you will come back to them in Chapter 1 of Book 2, and you will meet some examples throughout the module.

## Summary of Section 6.6

- The three types of RNA molecule – mRNA, tRNA and rRNA – are all essential for protein translation. mRNA carries the genetic code for synthesising a polypeptide, and tRNA is the key for deciphering the code; the amino acid specified by the code is brought to ribosome by tRNA and added to the growing polypeptide chain. rRNA is an essential component of the ribosome.
- Protein translation occurs in three stages: initiation, elongation and termination. All three stages take place at the ribosome, and are somewhat similar in prokaryotes and eukaryotes. A number of highly conserved translation factors facilitate these processes.
- In prokaryotes, translation initiation involves binding of the small ribosomal subunit to the ribosome binding site in the mRNA. This is followed by binding of the initiator tRNA ($tRNA_i^{Met}$) to the initiation codon (AUG) followed by binding of the large ribosomal subunit.
- In eukaryotic translation initiation, a preformed complex of the small ribosomal subunit, $tRNA_i^{Met}$ and specific translation initiation factors binds to the 5′ cap structure of the mRNA and scans along the mRNA to locate the initiation codon. Initiation is completed by binding of the large ribosomal subunit.
- During elongation, mRNA moves through the ribosome and as each codon enters, a tRNA with the correct anti-codon occupies the A (acceptor) site

of the ribosome. The ribosome forms a peptide bond between the amino acid carried by the tRNA in the A site and the amino acid already present in the P (polypeptide) site. The mRNA moves along and a new tRNA, with its attached amino acid, enters the A site (translocation).

- Translation terminates at a termination codon (UAA, UAG or UGA) when release factor (RF) protein enters the A site of the ribosome instead of a tRNA.
- Most eukaryotic proteins are modified post-translationally in a number of ways (including phosphorylation and glycosylation) before they become fully functional.
- Incompletely translated or misfolded proteins are degraded by the proteasome.

## 6.7 Final word

Many years of painstaking research and the development of sensitive techniques for detecting and quantifying RNAs and proteins have begun to reveal many of the complex processes that ensure cells can perform their different functions at the appropriate place and time. The level of expression of genes in prokaryotes and eukaryotes is controlled largely by the regulation of their transcription, but the processing and stability of eukaryotic mRNAs also contribute to the complexity of eukaryotic gene expression patterns. The translation of mRNA to synthesise proteins can also be controlled, and the protein itself may be modified to alter its activity. Although in this chapter the focus has been on the control of individual genes, it should be emphasised that in a cellular context it is the coordinated expression of the whole genome that determines the cell's development and function.

The next book in the module addresses 'the working cell', and considers in more detail the structure and function of cellular proteins; cell membranes and their role in transport of substances into and out of cells; how cells capture and use energy; how they transmit, receive and respond to signals from their environment; and finally, how they carry out a variety of different types of movement.

## 6.8 Learning outcomes

6.1 Describe the stages of gene expression in prokaryotic and eukaryotic cells.

6.2 Describe the important features of the nucleic acids and protein molecules involved in transcription and translation in prokaryotic and eukaryotic cells.

6.3 Distinguish between the mechanisms controlling the efficiency of transcription and translation in prokaryotic and eukaryotic cells.

6.4 Explain how post-transcriptional processing differs between prokaryotes and eukaryotes.

6.5 Outline some of the techniques used to study gene expression.

# Acknowledgements

Grateful acknowledgement is made to the following sources:

Figure 1.8: Adapted from Futuyma, D.J (2009) *Evolution*, 2nd edn, Sinauer Associates Inc.; Figure 1.11a: © Jarrod1 – Dreamstime.com; Figure 1.11b: Mark Grimson, Texas Tech University; Figure 1.11c: John Anderson/iStock photo.com; Figure 2.3: Source of image unknown; Figure 2.4: Dr Jack Bostrack, Visuals Unlimited Science Photo Library; Figures 2.5, 2.7c, 2.8, 2.18, 2.19, 2.20a, 2.21a, 2.22b, c, 2.23a and 3.24. Courtesy of Jill Saffrey, Department of Life, Health and Chemical Sciences, Open University; Figure 2.7a: Moll, R. et al. (2008) 'The human keratins: biology and pathology', *Histochemistry and Cell Biology*, vol. 129, with kind permission from Springer Science and Business Media, Springer Verlag; Figure 2.7b: Courtesy of Susan Van Noorden, Figure 2.9a Eye of Science Science Photo Library; Figure 2.9b: Steve Gschmeissner Science Photo Library; Figure 2.9c: Dr Richard Kessel and Dr Gene Shih, Visuals Unlimited Science Photo Library, Figure 2.9d Martin Oeggerli Science Photo Library, Figure 2.10a: Science VU, Visuals Unlimited Science Photo Library; Figure 2.10b: CNRI Science Photo Library; Figure 2.11 Prof. Heribert Cypionka Oldenburg University; Figure 2.12b: Astrid and Hanns-Frieder Michler Science Photo Library, Figures 2.13a and b Taken from the book *Freshwater Algae: Their microscopic world explored*, Canter-Lund and Lund (1995), Biopress, Figure 2.13c: Wim Van Egmond, Visuals Unlimited Science Photo Library; Figure 2.13d: Cristian Solari, University of Arizona; Figure 2.14b: Biospot Associates/Science Photo Library; Figure 2.15b: Thomas Deerinck, NCMIR/Science Photo Library; Figure 2.17b: Power and Syred Science Photo Library; Figure 3.2b Dr Gopal Murti/Science Photo Library; Figure 3.3: Courtesy of Heather Davies, Department of Life, Health and Chemical Sciences, Open University, Figure 3.10: Dr Torsten Wittmann Science Photo Library, Figure 3.11a, 3.17a, 3.19a and 3.22a: Courtesy of Professor Michael Stewart, Department of Life, Health and Chemical Sciences, Open University; Figure 3.12: Adapted from Purves, W.K. et al. (1998) *Life, the Science of Biology*, 5th edn, Sinauer Associates Inc; Figure 3.14: Dr Richard Kessel and Dr Gene Shih, Visuals Unlimited Science Photo Library; Figure 3.16: Photo printed with permission from O. L. Miller, B. A Hamkalo and C A Thomas Jr; Figure 3.21a: CNRI Science Photo Library; Figure 3.21b: Dr Don Fawcett Science Photo Library; Figure 3.22c: Dr Gopal Murti Science Photo Library, Figure 4.1a: National Institutes of Health, part of the United States Department of Health and Human Services; Figure 4.1c: Keith Bradnam, used under Creative Commons Attribution-ShareAlike License; Figure 4.1d: Nigel Cattlin Getty Images; Figure 4.1e: Janeff istockphoto.com, Figure 4.1f: Rama, licensed under the Creative Commons Attribution-Share Alike 2.0 France license; Figure 4.6b: Courtesy of the John Innes Foundation; Figure 4.12b: The American Philosophical Society; Figure 4.13: Dr Jeremy Burgess Science Photo Library, Figure 5.5: HHMI; Figures 5.8, 5.10 and 5.30: Alberts, B. et al. (2002) *Molecular Biology of the Cell*, 4th edn, Garland Science, a member of the Taylor and Francis Group; Figure 5.21: Shanel, used under a